GENETIC & MOLECULAR ANALYSIS OF HYBRID INCOMPATIBILITY IN DROSOPHILA

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Arundhati Shamoni Maheshwari

January 2012

GENETIC & MOLECULAR ANALYSIS OF HYBRID INCOMPATIBILITY IN DROSOPHILA

Arundhati Shamoni Maheshwari, Ph.D.

Cornell University 2012

Over evolutionary time genomes diverge by acquiring new mutations, some of which isolate species by erecting reproductive barriers. The study of such genes sheds light on how fundamental developmental processes diverge and result in new species. Hybrid incompatibility (HI) genes contribute to speciation by causing the sterility and inviability of interspecific offspring. The *Hmr* (*Hybrid male rescue*) and *Lhr* (*Lethal hybrid rescue*) genes are a major cause of hybrid lethality between *Drosophila melanogaster* and its sibling species *D. simulans*. Hybrid sons from this cross are normally inviable; however by mutating either the *D. melanogaster* ortholog of *Hmr* or the *D. simulans* ortholog of *Lhr*, viable adult hybrid sons are recovered. Like other HI genes, both *Lhr* and *Hmr* are rapidly evolving under selection.

In the first study, I asked whether the evolutionary history of selection for *Hmr* is confined to the hybridizing lineages. I conducted a population genetic survey of *Hmr* alleles from two sister species *D. yakuba* and *D. santomea*, whose common ancestor diverged from the common ancestor of *D. melanogaster* and *D. simulans* approximately 10 Myrs ago. I found that *Hmr* has diverged recurrently under positive selection in multiple independent speciation events, suggesting that *Hmr* is likely to be functionally diverging in multiple species.

In the second study, I examined the molecular nature of functional divergence for the HI gene, *Lhr*. Strikingly, I found that despite rapid evolution of

the *Lhr* coding sequence, hybrid lethal activity is not a derived function specific to one lineage, but instead a conserved function shared by both *Lhr* orthologs. Examination of the heterochromatic localization patterns of *Lhr* orthologs also failed to reveal any evidence of functional divergence. Instead I discovered that regulatory divergence underlies the asymmetric hybrid lethal activities of *Lhr* orthologs.

In the last study, I identified Lhr^2, a *D. simulans Lhr* allele with a highly unusual coding sequence, as a hybrid rescue mutation. I used it to identify a conserved region in the C-terminus of the LHR protein that is critical for hybrid incompatibility. Using the Lhr^2 allele I was able to assay the effect of an indel polymorphism that was segregating in the common ancestor of *D. melanogaster* and *D. simulans* on hybrid incompatibility. Notably, I found that this indel polymorphism contributes significantly to the functional divergence of *Lhr*.

BIOGRAPHICAL SKETCH

Shamoni Maheshwari was born and raised in New Delhi, India. She escaped schooling till the age of 13 and instead went to Mirambika, where she was encouraged to do whatever she genuinely felt like doing, even if that meant walking out of the classroom. From middle school onwards she was a student at Sardar Patel Vidyalaya, where she was first introduced to the idea of examinations and evaluations. Notwithstanding the initial amusement and indignation, she adjusted to this concept fairly well and graduated from high school in 1999. In the same year she joined the Biochemistry department at Sri Venkateshwara College at Delhi University. Three years later, she received a Bachelor of Science degree and was awarded a gold medal for graduating with the highest marks in her class. In 2002 she moved to the coastal metropolis of Mumbai (formerly known as Bombay) to pursue a Masters in Science at IIT Bombay. In 2004, she completed the masters at the top of her class and received a silver medal in recognition of her academic performance. A few months and thousands of miles later, she joined the Graduate Field of Biochemistry, Molecular and Cell Biology at Cornell University and in 2005 joined the lab of Dr. Daniel Barbash. There she immersed herself in the study of *Drosophila* speciation and was encouraged to explore it using all points of view be it, evolutionary, genetic, molecular or even biochemical. In August of 2011 she will finish her doctoral dissertation, after six years of more or less successfully crossing two different species of fruit flies.

To Papa and Nani.

ACKNOWLEDGEMENTS

To my adviser, Dan Barbash, I owe a huge debt of gratitude. His patience, support and unwavering belief in me were indispensable to the completion of this PhD. His scientific rigor and critical insight were key in realizing the research presented here. I was also fortunate to have an excellent committee, John Lis and Ken Kemphues, this thesis has significantly benefitted from their input. Additionally, Ken has been an invaluable counsel, both at a scientific and personal level. I thank him sincerely for our frank and open conversations and for his generosity with his time.

The vagaries of science would have been unbearable without having the company of my labmates to share them with. I thank you for all the debates and discussions, confidences and gossips, laughs and hysterics that we have shared for the past six years. I will miss you all dearly. I extend a special thanks to Nick Brideau and Heather Flores, who along with me were the first students to join the lab; I relied on your support tremendously. Without Patrick Ferree's mentoring and expertise in microscopy, an entire chapter in my thesis would not have been possible.

I would also like to acknowledge a number of faculty, colleagues and staff in the Molecular Biology & Genetics department that have contributed my success here. I often consulted Nick Fuda, Byron Williams and Raghuvir Vishwanatha when running into difficulty, and they always gave generously, both their time and expertise. I found sharing in the lively meetings of the Superfly community both enriching and scientifically rewarding. I received assistance from Xu Wang and Amanda Manfredo in pyrosequencing and Carol Bayles in microscopy. I have gone through three laptops and one desktop in my time in Cornell, so this thesis could literally not have been written without Mick Lipinto and his IT

staff.

Without my friends in Ithaca these past seven years would have been much more difficult and far less enjoyable. Specially my classmates Raga Krishnakumar, Nasun Ha, Xin Li, Raghuvir Vishwanatha and Sricharan Murugesan: for all the good times and for being there through the bad; and Greg Smaldone: for his help with the writing of this thesis and sitting through countless practice talks, but most importantly for keeping me sane.

Finally, I would not be where I am today without the unconditional love and support that I received from my parents. I thank them for their enduring faith in me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

THE GENETICS OF HYBRID INCOMPATIBILITIES[1]

1.1 Introduction

1.1.1 Defining hybrid incompatibilities

We suggest that hybrid incompatibility (HI) genes be defined as those that cause a measurable reduction in fitness in F1, F2 or BC1 generation interspecific hybrids (Figure 1.1). This definition allows for HI genes and their heterospecific interacting genes to be sampled when both heterozygous or homozygous. The utility of sampling later generation hybrids such as introgression genotypes is to provide increased resolution for mapping HI genes.We focus here largely on genes involved in hybrid sterility and lethality but also discuss some other kind of HI phenotypes. We largely examine HI between species, but also include some examples from inter-population hybrids. The boundaries between these classes of HI are not sharply defined, particularly for plants. We refer the reader to recent reviews about plant HIs and other aspects of HI genetics [1, 2, 3, 4].

1.1.2 The unique genetic properties of hybrids.

Hybrids are a unique genetic background that is not simply the sum of the two parental genomes. Consequently, the genetic properties of HI genes are often drastically different than one would predict from intraspecific genetic

[1]Written with equal contributions from S.M. & D.A.B and submitted to *Annual Reviews of Genetics*.

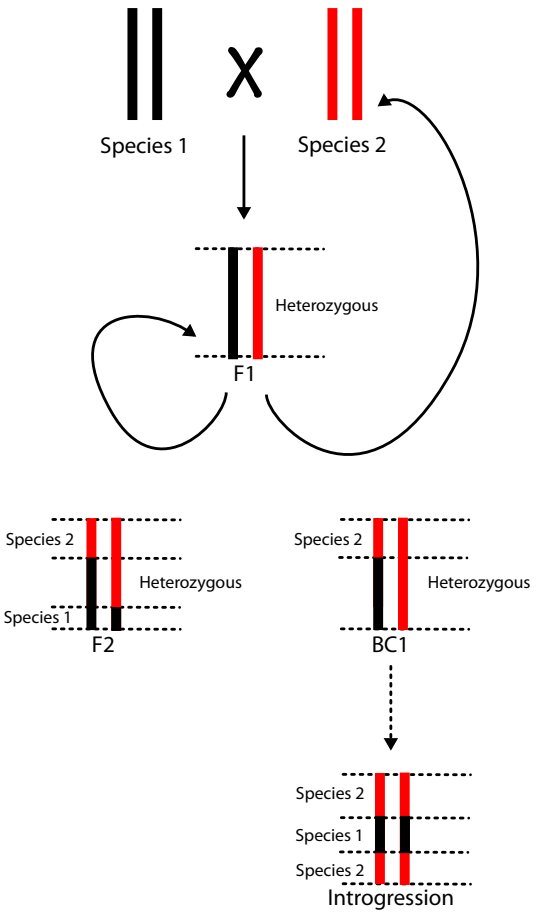

Figure 1.1: The genetic background of hybrids. Homologous chromosomes from two species are shown in red and black. The F1 hybrid has one allele from each parental species and is thus heterozygous at every locus (with the exception of the hemizygous sex chromosomes in the heterogametic sex). F2 progeny can have loci that are homozygous for each of the parental species as well as heterozygous loci. The BC progeny are similar to F2 progeny, except that they can only have loci homozygous for the backcross parent. Repeated generations of backcrossing result in an introgression genotype where only a single region is derived from the foreign species.

analyses. For example, mutations removing *Hmr* and *Ods* from pure-species *D. melanogaster* cause only mild reductions in fertility or viability [5, 6]. In contrast, the presence of these genes in hybrids causes complete lethality or sterility [7, 8].

1.1.3 Why are hybrids so different?

Diverse genetic mechanisms contribute to creating a unique genetic background in hybrids that is far from an additive combination of the two parental species.

Non-genic effects of DNA divergence. Sequence divergence between yeast species can directly cause HI, as described below. Non-specific euchromatic sequence divergence is the likely cause of the failure of somatic chromosome pairing in *Drosophila* hybrids [9]. Sibling species also often have large differences in heterochromatin content [10, 11]. We suggest that these large-scale DNA changes are under-explored as potential causes of aberrations in chromosome function and gene expression in hybrids.

Unpredictable epistatic interactions. As detailed below, HI phenotypes are caused by deleterious epistatic interactions between genes that have separately diverged along the two hybridizing lineages. By definition, these epistatic effects will be specific to the hybrid and therefore unpredictable from examining the parental species.

Gene expression changes. The net effect of genome-wide DNA divergence and unpredictable epistatic interactions, as well as species-specific divergence in regulatory pathways, can create large-scale aberrations in gene expression.

Studies discussed below comparing gene expression patterns in hybrids to their parental species have found many examples of non-additive gene expression in hybrids.

Failure of epigenetic silencing and imprinting. A variety of epigenetic processes control imprinting and silence repetitive DNAs, and failure of these epigenetic mechanisms can lead to the inappropriate activity of silenced genes and repetitive DNAs in hybrids (reviewed in [12]). These effects could in turn cause altered gene expression, DNA damage and chromosomal aberrations.

Asymmetric parental contributions. Uniparental inheritance of mitochondria, mRNAs, proteins and non-coding small RNAs from the maternal cytoplasm can create imbalances in hybrids between maternally inherited genetic material and the paternally inherited zygotic genome. Epigenetic processes are particularly susceptible to disruption in hybrids because they often depend critically on maternally inherited factors [12]. In one example, embryonic lethality induced by a paternally inherited block of satellite DNA, has been proposed to result from the inability of the maternal species to produce small RNAs required for its silencing [13].

1.1.4 Practical consequences of the unique hybrid genetic background.

If hybrids constitute a distinct genetic background that cannot be predicted from the parental species, then the genetic basis of HI must be studied in hybrids. We emphasize this apparent truism because it means that intraspecific genetics

may be a poor guide for understanding important questions such as why many HI genes are recessive in effect. It also suggests that efforts to reconstitute HI phenotypes within pure species may fail because the hybrid genetic background may itself be an essential factor required for the epistatic interactions among HI genes.

1.1.5 Why Mechanism Matters.

HI genes have been discovered in a wide range of taxa (Table 1), but the molecular mechanisms underlying hybrid breakdown are still largely unknown. We emphasize that intraspecific evolutionary forces drive divergence of HI genes, and HI phenotypes themselves are thus secondary consequences of these forces. Why then is it important to understand the mechanistic details of HI phenotypes? First, there may be key traits or developmental pathways that recurrently breakdown under the stress of divergent evolution, but that involve different genes in different lineages. Such a pattern will become more readily apparent by mechanistic investigations. Second, mechanistic understanding of incompatibilities can highlight the functions of a HI gene that have diverged between the hybridizing species, and thus suggest what evolutionary force drove the divergence. Third, we note below that different approaches to identifying HI genes make different assumptions about the mechanisms of HI gene function. Mechanistic studies will therefore help to assess which experimental approaches are most likely to discover HI genes. Fourth, the dominance properties of HI genes often cannot be directly determined from their genetic properties, because it is not always possible to construct heterozygous and homozygous genotypes in F1 and BC hybrids. One may instead need to infer dominance properties based

on mechanistic studies of HI gene action. Fifth, for cases where the molecular basis of the incompatibility is nongenic, for example epigenetic regulatory disruptions or a triggered mismatch repair pathway due to high sequence divergence, a mechanistic understanding may be the only method for revealing the cause of HI.

1.2 What Kind of Genetic Changes Cause HI?

1.2.1 Dobzhansky-Muller diverged genes.

We define Dobzhansky-Muller (D-M) genes as HI genes which cause incompatibility due to differences in their interaction patterns in hybrids versus pure species (Figure 1.2). We use "interaction pattern" in the most inclusive sense, including physical interactions of the HI gene's protein product, mRNA, *cis*-regulatory regions, or the DNA itself for noncoding HI genes. These differences could cause incompatibility either as the loss of an essential interaction or the gain of a negative ectopic interaction. We discuss D-M HI genes in detail below ("D-M HI genes").

Table 1: Diverged Dobzhansky-Muller genes that cause hybrid lethality or sterility.

Locus	Species Cross	HI phenotype	Approach used for identification	No. of loci involved	Molecular function in pure species	Mechanism of Incompatibility	Proposed HI-causing divergence	Proposed evolutionary basis	Refs
AEP2	*Saccharomyces cerevisiae X S. bayanus*	F2 sterility	Chromosome substitution; screening genomic library for complementation	2;AEP2 & OLI1	Nuclear-encoded mitochondrial translation factor	L.o.f. *S. bayanus* Aep2 fails to translate *S. cerevisiae* OLI1 mRNA	Changes in CDS	Mutational, response to high mitochondrial mutation rate or Ecological, adaptation to novel carbon sources	[43, 94]
OLI1	*S. cerevisiae X S. bayanus*	F2 sterility	Candidate gene testing		F0-ATP synthase subunit		Divergent 5'-UTR sequence		[43,94]
MRS1	*S. cerevisiae X S. bayanus or S. paradoxus*	F2 lethality	Selection for respiration deficient hybrid clones; screening genomic library for complementation	2; MRS1 & COX1	Nuclear-encoded mitochondrial splicing factor	L.o.f. *S. cerevisiae* Mrs1 fails to splice ancestral COX1 mRNA	Three amino acid substitutions		[78,94]
COX1	*S. cerevisiae X S. bayanus or S. paradoxus*	F2 lethality	Candidate gene testing		Subunit of cytochrome c oxidase		Loss of intron in *S. cerevisiae*		[78,94]
AIM22	*S. cerevisiae X S. bayanus*	F2 lethality	Selection for respiration deficient hybrid clones; screening genomic library for complementation	Unknown	Nuclear-encoded mitochondrial lipoate protein ligase	Unknown	Unknown		[78, 94]
Lhr	*Drosophila melanogaster X D. simulans*	F1 male lethality	Mapping of rescue allele; transgenic complementation	$\geq$ 3; Lhr & Hmr & ?	Heterochromatin-associated protein of unknown function	G.o.f. Determined by genetic assays	Divergent *cis*-regulation and rapid evolution of CDS	Mutational, heterochromatin turnover, or Conflict, involving heterochromatin-mediated distortion or drive	[51, 182, this thesis]
Hmr	*D. melanogaster X D. simulans*	F1 male lethality; F1 female lethality and sterility	Mapping of rescue allele; transgenic complementation		Heterochromatin-associated protein of unknown function		Rapid evolution of CDS		[7, 50, 127]
Zhr	*D. simulans X D. melanogaster*	F1 female lethality	Mapping of rescue allele; cytological confirmation	Unknown	Pericentric satellite DNA	G.o.f. Missegregation of satellite DNA	Satellite DNA unique to D. melanogaster		[111]
OdsH	*D. simulans X D. mauritiana*	BC-like male sterility	Introgression mapping; transgenic complementation	Unknown	Heterochromatin-associated protein of unknown function	G.o.f. Ectopic localization to the *D. simulans* Y	Rapid evolution of DNA-binding homeobox domain	Conflict, suppression of heterochromatin-mediated distortion or drive	[8, 85]
Ovd	*D. pseudoobscura bogatana X D. pseudoobscura pseudoobscura*	F1 male sterility	Introgression mapping; transgenic complementation	$\geq$3; Ovd & ?	Unknown	Unknown	Six amino acid substitutions	Conflict, suppression of sex-ratio distortion	[83, 84]
Nup96	*D. melanogaster X D. simulans*	BC-like male lethality	Mapping of incompatible hemizygous region	$\geq$3; Nup96 & ?	Subunit of nucleopore	L.o.f. and/or G.o.f.	Rapid evolution of CDS	Ecological, response to pathogens, or Conflict, suppression of segregation distorters	[88, 90, 107]
Nup160	*D. melanogaster X D. simulans*	BC-like male lethality & female sterility	Introgression and hemizygosity mapping; transgenic complementation	$\geq$3; Nup160 & ?	Subunit of nucleopore	L.o.f. and/or G.o.f.	Rapid evolution of CDS		[57, 89]
Xmrk-2	*Xiphophorus maculatus X X. helleri*	F2 lethality	Mapping in backcrosses; transgenic induction of phenotype	Unknown	Receptor tyrosine-kinase	L.o.f. Deregulated expression of Xmrk-2	Novel promoter region acquired after gene duplication	Unknown	[125]
Prdm9	*Mus musculus musculus X M. musculus domesticus*	F1 male sterility	Mapping of HI polymorphism; transgenic complementation	$\geq$3; Prdm9 & Hst^{ws} & ?	Determinant of meiotic recombination hotspots	L.o.f. Transgenic addition of the *Prdm9* fertile allele to the incompatible background rescues sterility	Rapid evolution of DNA-binding zinc finger domain	Mutational, response to meiotic hotspot turnover	[48, 97]
S5	*Oryza sativa indica X O. sativa joponica*	F1 female sterility	Mapping of HI polymorphism; transgenic complementation	2; *O. s. indica* S5 & *japonica* S5	Proteolytic enzyme in the cell wall	G.o.f. Dominant negative interaction between S5 alleles	Two amino acid substitutions	Mutational, inferred from minimal changes in gene structure	[76]
SaM	*O. sativa indica X O. sativa joponica*	F1 male sterility	Introgression mapping; transgenic complementation	3; *O. s. japonica* SaM & *O. s. indica* SaF & SaM	Predicted SUMO E3 ligase-like protein	G.o.f. Truncated SaM creates negative interaction with SaF	Change in intron structure		[77]
SaF	*O. sativa indica X O. sativa joponica*	F1 male sterility			F-box containing protein of unknown function		A single amino acid substitution		[77]

1.2.2 Chromosome rearrangements/karyotypic changes.

If two hybridizing species have different karyotypes then hybrids may suffer from meiotic defects or produce gametes that are aneuploid, and thus be sterile (reviewed in [2]). Polyploid hybrids are frequently sterile and are common among plants [14]. Chromosomal inversions have also been widely implicated in causing hybrid sterility. Recent work, however, suggests that sterility is not directly caused by the inversions interfering with meiosis, but rather because inversions are more likely to harbor diverged D-M HI genes compared to non-inverted regions [15, 16, 17].

1.2.3 Gene transposition/reciprocal gene loss.

The simplest possible cause of HI is if hybrids lack a copy of an essential gene. We classify examples of such genes as distinct from D-M genes because their absence causes HI and does not require any functional divergence or change in the interaction patterns of the gene. Gene absence, however, is not contradictory to the widely-accepted Dobzhansky-Muller model [18, 19].

Gene transposition is one mechanism that can lead to this condition [20]. If a gene transposes from one chromosome to another, then random segregation in a BC1, F2 or later generation will lead to hybrids with no copies of the gene. An example is *JYalpha*, which transposed from chromosome 4 to chromosome 3 in the ancestor of *D. simulans*. A homozygous introgression of *D. simulans* chromosome 4 into *D. melanogaster* thus lacks *JYalpha* and is male-sterile because this gene is required for male fertility [21].

Force and Lynch have argued that the asymmetric resolution of gene duplicates may be a potent cause of HI [18]. Degeneration after gene duplication can result in different paralogs becoming the sole active copy in different species, a situation similar to gene transposition. This process has been shown to cause lethality or sterility in inter-strain hybrids of Arabidopsis [22] and rice [23, 24].

Whole-genome sequence data allow accurate estimations of the potential role of these processes in causing HI. Single-gene transposition is unlikely to play a major role in causing HI in Drosophila, because the number of estimated hybrid male sterility loci between closely related species [25, 26, 27, 28] vastly exceeds the estimated number of high-confidence gene transpositions [29].

Scannel et al. [30] have reconstructed the ancestral pre-whole-genome duplication (WGD) genotype in yeast to argue that rapid loss of gene duplicates after WGD occurred on the lineage leading to *S. cerevisiae*, at a rate that may have contributed to speciation between post-WGD species. On the other hand, genome-scale analysis of 5 fish species that post-date the teleost-specific WGD found relatively few candidates of asymmetric paralog loss that were potentially causal for speciation [31].

Plants might be expected to be more susceptible to this cause of HI because they often contain large families of paralogous genes and higher rates of WGD and polyploidy than animals [14, 32]. The *HPA* paralogs in Arabidopsis and the *DPL1/DPL2* paralogs in rice implicated in HI have degenerated multiple times independently in their host species, suggesting that the same genes might recurrently mutate to cause HI [22, 23]. Mizuta et al. [23] conclude, however, that reciprocal gene silencing after duplication can only account for 1 out of about 30 known reproductive barriers between rice strains, and they note similar propor-

tions in Arabidopsis. It appears then that while certain paralogous genes may be particularly susceptible to reciprocal gene silencing, the overall rate is rather low.

1.2.4 Sequence divergence.

Sequence divergence between parental genomes will lead to DNA mismatches during the strand invasion step of meiotic recombination. This can in turn activate the mismatch repair pathway and block recombination. Meiotic crossovers are necessary for accurate chromosome segregation, so a reduction in recombination increases the rate of aneuploidy which in turn can cause sterility. Deleting mismatch repair proteins can partially suppress yeast hybrid sterility, arguing that sequence divergence is a direct cause of HI in yeasts [33, 34].

1.2.5 Transposable elements and non-coding repeats.

Eukaryotic genomes vary greatly in size, largely due to differential abundance of non-coding satellite DNAs and transposable elements. *Zhr* is the only known example of a non-coding DNA that directly causes HI (Table 1). There are also multiple reports of increased TE activity in hybrids (reviewed in [2, 12]) although the effects on hybrid fitness are not clear. However, there is now a resurgent interest in the possible role of non-coding DNAs in HIs due to an increased understanding of the molecular mechanisms that suppress selfish DNAs. As we discuss below under "internal genetic conflicts", selfish DNAs exert strong selective pressures on their host species, which can lead to rapid evolution of

host suppressor genes that in turn cause HI.

1.2.6 Dosage imbalances

Dosage-sensitive incompatibility is best illustrated by interploidy crosses between *Arabidopsis* species [35]. When diploid *A. thaliana* ovules are fertilized by pollen from diploid *A. arenosa*, only 5% of hybrid seeds are viable. However, crosses using tetraploid *A. thaliana* ovules and diploid *A. arenosa* pollen (4x X 2x) produce 79% viable hybrid seed. Interestingly, expression of ATHILA retrotransposons is strongly induced in hybrid seeds, but the degree of deregulation is lower in the more compatible interploidy cross (4x X 2x). Joseffson et al. [35] determined that the derepressed ATHILA retrotransposons are strictly paternally encoded. The authors suggest that hybrid inviability is due to insufficient dosage of *A. thaliana* maternal repressors relative to the number of ATHILA retrotransposon targets in the paternal *A. arenosa* genome. An increase in the maternal ploidy level rescues seed lethality by correcting the imbalance between the number of maternally expressed repressive factors and paternally inherited transposons.

1.3 Approaches for Finding HI Genes

1.3.1 Mapping in backcrosses.

Standard genetic mapping can be used if one sex of F1 hybrids is fertile. For example, F1 hybrid females of *D. pseudoobscura* and *D. persimilis* are viable and

fertile, but the hybrid males are sterile. By backcrossing F1 females to pure species males, recombinant backcross males can be phenotyped for fertility and genotyped using visible and molecular markers. This approach was pioneered by Dobzhansky [36], who concluded that hybrid male sterility factors are not evenly distributed among chromosomes and also that sterility effects are additive.

The major advantage of this method is that it can assay incompatibilities in early generation hybrids such as the BC1 or F2 generations which are most relevant to reproductive isolation in the wild. In particular, it can capture complex multi-locus interactions. A second advantage is its potential to be easily extended using a high density of molecular markers. For example, quantitative trait loci (QTL) mapping has been used to estimate the number and location of regions contributing to hybrid male sterility along entire chromosomes or across the genome [37, 38].

The major limitation is that the mapping resolution will be relatively low in the absence of high recombination rates and large numbers of hybrids to assay. And like most other approaches that use recombination for mapping, genes residing within regions inverted between the hybridizing species cannot be localized by backcross mapping.

1.3.2 Deviations from Mendelian ratios.

An alternative method is to look for deviations from Mendelian inheritance ratios of the parental alleles in backcross or F2 populations. Compared to QTL mapping which is typically performed on a predefined trait, this method has

the advantage of potentially discovering different types of reproductive barriers including gametophytic sterility and segregation distortion occurring in the F1 parents as well as zygotic lethality in the F2 or BC progeny. The method has been most widely applied to seed-bearing plants, for example refs. [39, 40] but also to ferns [41] and *Nasonia* wasps [42].

1.3.3 Introgression.

Dobzhansky [36] described a natural extension to mapping in F2 or backcross populations, namely to repeatedly backcross a candidate HI region from the foreign species into the hybridizing species, until the region has been introgressed into an otherwise pure species background. Introgressions can then be whittled down by recombination in order to map the HI genes. This approach was used to clone the *Odysseus* hybrid sterility gene [8]. Introgressions can also include whole chromosomes, as has been done in yeast and mice [43, 44].

One advantage of the introgression approach is that recombination is done within the host species, which is generally easier to achieve with larger numbers compared to mapping in backcross hybrids. A second advantage is that introgressions can be performed genome-wide to address rates and patterns of HIs such as between different chromosomes and between males and females [26, 45, 27, 28].

The major disadvantage of introgressions is that they will only detect relatively simple interactions between the introgressed region and the host genome (Figure 1.1). Incompatibilities caused by interactions between unlinked genes of the donor species will not be detected unless multiple introgressions are com-

bined together.

1.3.4 Polymorphic HI factors.

One advantage of mapping HI genes still segregating in the population is that it may bias for the recovery of incompatibilities that are relatively recent. Many HI phenotypes in plants show great variability in penetrance depending on the strains used. Some strains of *Crepis tectorum* produce lethal F1 hybrids when crossed to *C. capillaris*. Hollingshead [46] discovered a strain of *C. tectorum* that produced lethality of only half the hybrids, and deduced that this strain was heterozygous for a single hybrid lethality factor. She also made the important discovery that this lethal factor causes no phenotype within *C. tectorum* even when homozygous.

More recently the first vertebrate HI gene, *Prdm9*, was mapped using this approach. The house mouse subspecies, *Mus. m. domesticus* and *M. m. musculus*, diverged from a common ancestor only within the last 0.5 million years yet are isolated by at least two sets of D-M incompatibilities [47]. The incompatibility system is very complex, with mice from both species polymorphic for the sterility-causing loci. Some laboratory inbred strains of *M. m. domesticus*, such as C57BL10, when crossed to wild mice from *M. m. musculus* produce completely sterile F1 hybrid sons, while other strains, such as C3H/Di produce fertile progeny. This strain-specific difference was mapped to a single polymorphic locus on chromosome 17, *Hybrid sterility1* (*Hst1*), and shown to correspond to the gene *Prdm9* [48]. Interestingly, like *Hst1*, both sterile and fertile alleles of the *Hst1* interacting locus Hst^{w} are segregating in multiple populations of wild

mice [49].

1.3.5 Rescue or suppressor alleles.

F1 hybrid sons or daughters of *D. melanogaster* and *D. simulans* are lethal, depending on the direction of crossing. Several rare alleles have been found in laboratory or wild strains that produce viable hybrids of the otherwise lethal sex. These suppressor strains led to the identification of the *Hmr*, *Lhr* and *Zhr* HI loci (Table 1). The ease of mapping these suppressors relative to non-rescuing wild type alleles is one advantage of this approach. This approach could also potentially be used to find all genes causing an incompatibility by laboratory mutagenesis. However, one concern with rescue alleles is that they might act as suppressors of HI while not being directly involved themselves, for example by creating neomorphic alleles. To date this has not been a problem, as the *Hmr*, *Lhr* and *Zhr* rescue alleles are loss-of-function mutations of wild type genes that directly cause hybrid lethality [50, 51, 52, 53]. A more severe limitation is that this approach will likely only work for HIs that have a simple genetic basis, because complex multilocus interactions are unlikely to be suppressed by a single gene mutation.

1.3.6 Other genetic screens.

Laboratory model organisms such as *D. melanogaster* have collections of strains that delete defined segments of the genome. One can then cross such deletions into hybrids and look for regions that when hemizygous cause HI; the interpre-

tation of such effects being that the deletions are uncovering recessive HI genes. Several such screens have been performed [54, 55, 56], and led to the discovery of two adaptively evolving nucleopore subunit genes, *Nup96* and *Nup160*, that cause hybrid lethality (Table 1).

One concern with this approach is that the screens are based on fitness of F1 hybrids that are heterozygous for large deletions, which is a genotype never found in normal hybrids. Lethality could potentially be caused by haploinsufficiency rather than a true hybrid incompatibility. Hybrids are more sensitive to imbalances of gene dosage than pure species, making it likely that adverse effects of haploinsufficiencies are exacerbated in the hybrid background. These effects may be particularly problematic when trying to make quantitative estimates of HI gene density, such as in ref. [56]. One way to confirm that a hemizygous deletion is uncovering bona fide D-M HI genes is to ask whether the same region also causes HI when made homozygous in hybrids. Such a test has been performed for the *Nup160* region by introgressing it from *D. simulans* into *D. melanogaster*, where it was found to cause lethality to F1 hybrid males when homozygous [57].

1.3.7 Transcriptional profiling.

Both intuition and modeling suggest that disruptions of gene regulatory pathways may contribute to HIs [58]. Several studies have used whole-genome expression profiling to reveal substantial differences in gene expression between hybrids and their parental species [59, 60, 61, 62, 63, 64]. As noted by Michalak and Noor [63], the challenge then becomes determining which changes are asso-

ciated with HI and identifying the genetic events causing gene misexpression. The results to date are mixed. A study comparing lethal *D. melanogaster*/*D. simulans* hybrid males with those rescued by the *Hmr* mutation found relatively few genes differentially expressed. Proteosome subunit genes were overrepresented among these, but limited genetic tests failed to find any effects on hybrid fitness [65]. Walia et al [66] compared viable and inviable seeds from interspecific Arabidopsis hybrids, identified a cluster of coregulated AGAMOUS-LIKE (AGL) genes as differentially expressed, and found that mutations in two of them significantly increase hybrid seed viability. The authors suggest that epigenetic dysregulation of AGL genes is likely the downstream effect of failure or incompatibility in hybrids of the polycomb-repressive complex (PRC).

An alternative approach put *D. simulans*/*D. mauritiana* hybrids through 5 generations of backcrossing, and looked for associations among gene expression levels, fertility and genotype [67]. Several genes were consistently underexpressed in sterile hybrids, and this phenotype was associated with a region containing the *OdsH* hybrid sterility gene. Finer-scale mapping of this region could potentially identify the specific gene(s) that correlate with altered gene expression. Although gene expression profiling has not led to the discovery of new HI genes, it has revealed interesting patterns in hybrids such as differences in gene expression between the sexes or between X chromosomes and autosomes [68, 69, 70].

1.3.8 Biochemical and molecular tests.

Hybrids between isolated populations of the copepod *Tigriopus californicus* show a range of hybrid breakdown phenotypes in the F2 generation that suggest incompatibility between the nuclear and mitochondrial genomes. Sensitive *in vitro* biochemical assays using preparations of cytochrome c (CYC) variants show a reduced rate of oxidation when tested with mitochondrial extracts from different populations, and led to the proposal that nuclear-encoded *CYC* genes are coevolving with mitochrondrial-encoded cytochrome c oxidase (*COX*) genes within *T. californicus* populations [71]. Genetic studies suggest that additional incompatibilities occur between other mitochondrial and nuclear genes involved in the mitochondrial electron transport system [72].

1.4 Genetic Interactions and the D-M Model

How do genes causing deleterious phenotypes in hybrids evolve without species themselves going through reduced fitness? The solution is that each of a pair of interacting genes evolves independently in separate lineages, and the deleterious interactions between them only occur in hybrids (Figure 1.2). We refer to this as the Dobzhansky-Muller model [73, 74], although Orr has convincingly argued that Bateson independently derived the same solution [75]. The simplicity and generality of the D-M model explains why it has continued to guide HI genetics for over 70 years. Here we discuss several important open questions that delve deeper into the implications of this model.

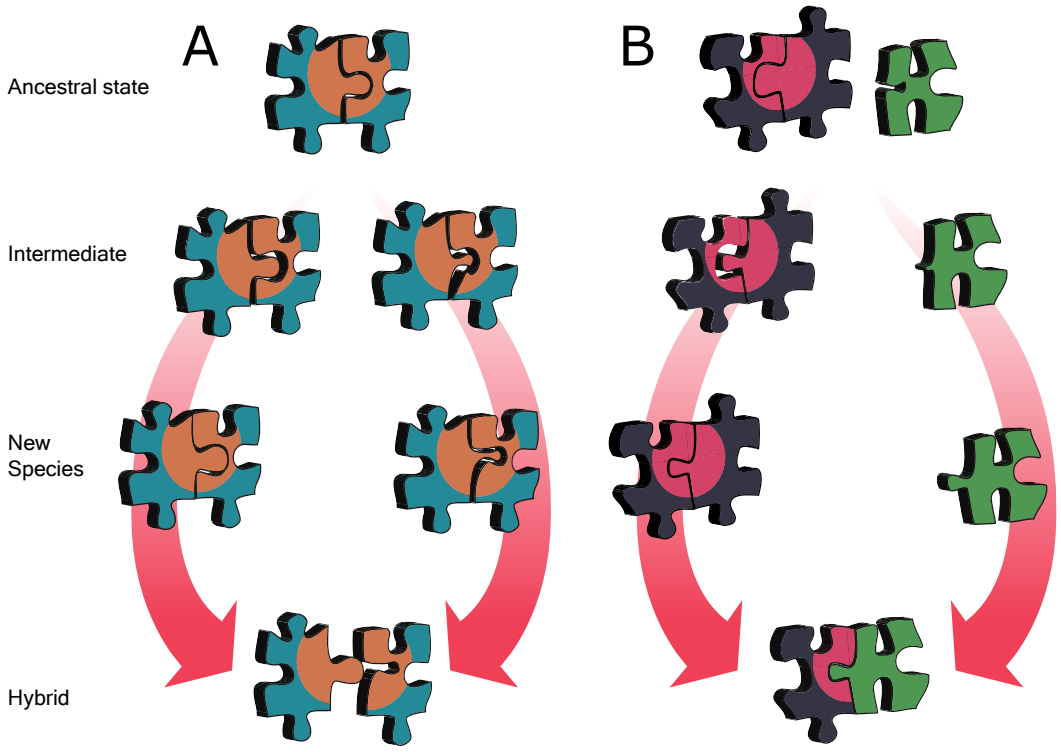

Figure 1.2: Dobzhansky-Muller incompatibility models.(A) Two-locus loss-of-function incompatibility. In an ancestor (top) two interacting loci are depicted. Over time, mutations arise leading to lineage-specific co-evolution between the interaction partners. In the hybrid (bottom), the diverged alleles are now incapable of interacting. Hybrid incompatibility results from the loss of this essential interaction. (B) Two-locus gain-of-function incompatibility. Three loci are depicted, two of which interact in the ancestor while the third one (in green) does not. In one of the lineages, the interacting partners co-evolve in the genetic background of the ancestral third loci. In the second lineage, mutations arise leading to divergence of the non-interacting third locus. In the hybrid, this lineage-specific divergence results in the gain of an ectopic interaction.

1.4.1 Are HIs simple or complex?

The D-M model is typically illustrated as involving 2 loci but it can be easily extended to multiple loci [19]. The strongest experimental proof that a two locus interaction is causing HI is if by introducing the D-M partner into the hybridizing species, the incompatibility phenotype is recreated. This has been achieved for the inter-allelic hybrid sterility locus *S5*, where rice plants of the *japonica* strain when transformed with *indica S5* allele have greatly reduced fertility [76]. A more complex three-locus incompatibility also occurs in rice, involving both the *indica* and *japonica* alleles of the *SaM* locus and the *indica* allele of the adjacent *SaF* locus, and has been similarly recreated in a pure species background [77]. The yeast mitochondrial-nuclear incompatibilities involving *AEP2* and *OLI1*, and *COX1* and *MRS1* also each appear to be fully explained by a pairwise interaction [78, 43]. Numerous other examples of putative 2-locus HIs have been discovered in plants although the causal genes have not yet been identified, for example refs. [79, 46, 80].

The question of whether HIs are genetically complex is often framed in terms of how many total genes contribute to a broadly defined HI phenotype, such as male sterility. For some *Drosophila* species, the answer from introgression studies is clearly "very many" [26, 81, 27]. But it is unclear whether these incompatibilities represent many evolutionary independent two-locus HIs caused by distinct mechanisms, a small number of multi-locus HIs with the same mechanistic basis, or something in between. Several approaches suggest themselves for addressing the complexity of HI phenotypes. One is to fully dissect the effect of a single introgression. One such high resolution study found that full sterility requires the interaction of multiple loci [82]. A second is to combine together in-

trogressions that have weak phenotypes in order to detect interactions between regions causing more severe HI phenotypes [9]. A third approach has used QTL mapping to demonstrate that full sterility typically requires epistatic interactions among multiple loci [38]. Finally, studying more closely related species offers the promise of completely dissecting an HI phenotype. Male sterility between subspecies of *D. pseudoobscura* offers a strong opportunity to understand fully the epistatic interactions causing HI. While additional regions may be involved, 4 interacting loci have been mapped and one of them, *Ovd*, has already been identified [83, 84].

1.4.2 Is incompatibility in hybrids caused by loss or gain of gene function?

Often, even after the molecular identification of the HI loci the mechanism underlying the incompatibility phenotype is unclear. The Dobzhansky-Muller model is equally consistent with HI being caused by either the loss of an essential interaction or the gain of a negative ectopic interaction. However, a discussion of HI loci using standard genetic terminology, such as gain or loss of function alleles, is complicated by divergence along the two lineages. Genetic interactions unique to the hybrid may evolve, such that the function in question may not necessarily be predictable from pure-species effects. Additionally, the two orthologs of a HI gene have evolved independently, thus each ortholog of an HI gene is unlikely to cause the same effects. There is also a practical problem, namely that one cannot always perform the necessary formal genetic tests that require manipulating HI gene activity, most critically introducing defined

loss-of-function mutant alleles. Mechanistic investigations, however, often provide interpretable data that can distinguish between loss and gain of function mechanisms.

Loss-of-function HI genes. Loss of an essential interaction in the hybrid background can evolve as the by-product of lineage-specific coevolution of interaction partners (Figure 1.2). Chou et al. [78] found that the *S. cerevisiae* nuclear-encoded mitochondrial splicing factor MRS1 is incompatible with the *S. bayanus* or *S. paradoxus* mitochondrial-encoded *COX1* gene. Replacement of *S. bayanus* or *S. paradoxus MRS1* with *S. cerevisiae MRS1* results in drastic reduction of mature mitochondrial *COX1* mRNA. Comparison of the *COX1* gene structure revealed that an evolutionary conserved intron is missing only in the *S. cerevisiae* lineage. This observation strongly suggests that coevolution between *COX1* intron structure and its splicing factor MRS1 led to the loss of the ancestral splicing activity in *S. cerevisiae MRS1*. This hypothesis is further supported by functional analysis ascribing the loss of splicing activity to three amino-acid substitutions derived in the *S. cerevisiae* lineage. In another example of cytonuclear incompatibility, a sporulation defect is caused by the inability of the Aep2 protein from *S. bayanus* to regulate the translation of the *S. cerevisiae OLI1* mRNA [43]. *AEP2* therefore also behaves as a loss-of-function mutation in a hybrid background.

Gain-of-function HI genes. In a gain of function scenario the HI gene product either acquires a negative ectopic interaction with a gene in the hybrid background that does not occur in pure species or the hybrid genome becomes sensitive to gene activity in a way that is deleterious (Figure 1.2).

For the hybrid male sterility gene *OdsH* molecular evidence strongly suggests gain of ectopic localization as the basis of incompatibility [85]. *D. mauri-*

tiana OdsH when expressed in *D. simulans* exhibits extensive mislocalization to the *D. simulans* Y chromosome, which is likely caused by the dramatic sequence divergence in the *OdsH* DNA-binding homeodomain.

Hybrid sons from a cross of wild type *D. melanogaster* mothers to *D. simulans* fathers are inviable. A loss of function mutation at the *D. simulans Lhr* locus reverses this lethality, demonstrating that the presence of wild type *Lhr* causes dominant lethality to hybrids [51]. However, hybrid lethality is not simply the consequence of an unregulated excess of *Lhr* activity, because ubiquitous over-expression of *Lhr* within *D. melanogaster* has no effect. Thus, incompatibility is best explained by the Lhr protein acquiring a negative ectopic interaction in the hybrid background, but the molecular nature of this interaction remains to be identified. Unlike *OdsH*, *Lhr* orthologs do not appear to mislocalize when expressed in a foreign species [86].

Three gain-of-function incompatibility systems have been recently identified in rice [76, 77, 87]. In each case a gain-of-function mechanism is inferred because addition of the HI gene ortholog into a wild type background of the hybridizing species recreates the incompatibility phenotype.

Complex cases. The *Nup96* and *Nup160* nucleoporins, encoding subunits of the Nup107 nucleopore-subcomplex, cause HI between *Drosophila melanogaster* and its close relative *D. simulans* [88, 89]. They and other members of this complex show rapid sequence divergence driven by adaptive evolution, suggesting functional divergence due to lineage-specific coevolution of the subcomplex [90]. A simple loss-of-function scenario would mean that a functional nucleopore fails to assemble when only *D. simulans* subunits of *Nup96* or *Nup160* are present. An alternative scenario is that the lineage-specific divergence could be acting

as deleterious gain of function mutations, for example by preventing other subunits from assembling or by poisoning nucleopore function. A realistic interpretation is that these HI genes may share both loss- and gain-of-function attributes [91].

1.5 What Evolutionary Forces Lead to HIs?

The D-M model proposes that HI is a secondary consequence of intraspecific divergence. The answer to the evolutionary origins of HI therefore lies in understanding the evolutionary origins of the culpable divergence. We consider here 3 major evolutionary mechanisms causing this divergence: ecologic adaptation, neutral mutational processes, and internal genetic conflicts.

We highlight in this section two important points. One is that all three of these evolutionary mechanisms can lead to signatures of positive selection in HI genes. While the terms "positive selection" and "adaptive evolution" are often used interchangeably, signatures of selection in a gene can also arise in response to non-adaptive neutral mutational processes or internal genetic conflicts. Both processes can induce negative fitness costs, to which the organism adapts by evolving suppressors, but the "adaptation" is to ameliorate negative fitness costs. We therefore suggest that it is premature to classify the evolution of HI genes *Hmr* and *Nup96* as reflecting the action of "ecological factors" simply because they have evolved under positive selection [92].

A second important point is that distinguishing among these 3 very different mechanisms is challenging. We will illustrate this by discussing how all 3 mechanisms could plausibly explain the origin of nuclear-mitochondrial incom-

patibilities.

1.5.1 Ecological Adaptation.

The term ecological adaptation of HI genes follows from the widely used (though not always appreciated) term "ecological speciation" [93, 92]. This type of HI would be a consequence of organisms adapting to new or altered environments, food sources or external pathogens.

Chou and Leu [94] note that nuclear-mitochondrial yeast incompatibilities could be a consequence of yeast species adapting to different carbon sources, reflecting evolved differences in the ability of species to utilize non-fermentable carbon sources. Since mitochondria are a critical organelle for regulating energy metabolism, this type of ecological differentiation is a plausible driver of divergence of mitochondrial genes.

The widespread observation of necrosis among both intraspecific and interspecific plant hybrids is the most plausible example of selective pressures driven by pathogens as the cause of HI. A number of plant immune system genes and gene families are evolving rapidly and some have been implicated in hybrid necrosis [1].

1.5.2 Mutational Processes.

The cases of chromosome re-arrangements, gene transpositions, reciprocal gene silencing, and sequence divergence described in the section "What kind of ge-

netic changes cause HI" are clear examples of how mutational processes can directly cause HI.

The evolutionary pattern of nuclear-mitochondrial incompatibilities might also be explained by the high mutation rate of the mitochondrial genome due to a high level of oxidative damage, an inefficient DNA repair system, and the lack of protective histone molecules (reviewed in [94]). In this scenario, the mitochondrial-encoded HI genes evolve rapidly for mutational reasons, and their nuclear-encoded interacting partners co-evolve within lineages to remain compatible. In the hybrid, mismatched partners would fail to interact and thus cause HI.

Similar to high mutation rates in the mitochondrial genome, purely mutational changes in the nuclear genome can also alter the internal genetic environment of an organism sufficiently to be the driver of positive selection in protein-coding genes. One example may be the hybrid sterility gene *Prdm9*. Recombination hot-spots are sites in the genome with a high frequency of meiotic recombination. Being preferential targets of recombination they preferentially accumulate double-strand breaks (DSBs). To promote chiasmata formation, meiotic DSB repair inherently favors the non-sister chromatid as template. If the non-sister chromatid is polymorphic for the recombination hot-spot sequence, its sequence will "erase" the initiating hot-spot sequence. The survival of recombination hot-spot sequences has therefore been a paradox: How do sequences that by definition encode their own extinction persist [95]? The rapid evolution of the zinc fingers of *Prdm9* that encode its hot-spot recognition capacity provides a clever answer to this long-standing question [96]. As established meiotic hot-spots degrade, the zinc fingers encoded by *Prdm9* will evolve to recognize new

sequences, thus "reinventing" a new hot-spot motif. As this process repeats itself *Prdm9* will accumulate signatures of adaptive evolution, as has been detected among mammalian species [97]. The interesting discovery that allelic variation in the human *Prdm9* zinc finger array contributes to hot-spot activity and meiotic instability further supports the hypothesis of co-evolution between *Prdm9* and meiotic hot-spot sequences [98].

The mutational process of unequal crossing over can be the origin of HI-causing divergence in non-coding satellite DNAs [99]. Small-scale variation in satellite abundance is likely to be neutral in effect, but over time larger scale differences may accumulate and become deleterious, for example by interfering with DNA replication or chromosome segregation. Host genes would then evolve to ameliorate these fitness costs. HI genes encoding heterochromatin proteins such as *Lhr* and *Hmr* potentially fit this scenario.

1.5.3 Internal Genetic Conflicts.

Defining internal genetic conflict. Genetic conflict between two entities requires that a selective advantage to one comes at a cost to the other. Without cost there is no conflict. Proving cost is therefore key to showing that two genetic entities indeed have conflicting interests. Conflict occurs at multiple levels. One kind of conflict is external, for example between microbial or viral pathogens and their host species. Because it involves an organism adapting to external factors we classify this type of situation above under ecological adaptation.

Internal conflicts can occur between the nuclear genome and the cytoplasmically-transmitted mitochondrial and chloroplast genomes [100]. Because these or-

ganelle genomes are exclusively transmitted through females, conflicts with the nuclear genome will preferentially harm males. This type of conflict may explain the high prevalence of cytoplasmic male sterility in plants. The resolution of conflicts between mitochondrial and nuclear genomes can lead to co-evolution of their genomes, with HI resulting from disruptions of co-adapted mitochondrial and nuclear gene complexes.

Internal conflicts will also occur when an organism is challenged by selfish genetic elements within its own genome. Transposable elements are a prime example, because they are self-replicating genomic parasites that can cause mutations and chromosomal rearrangements in the host genome [101]. A second type of internal conflict is driven by heterochromatic repeats that can lead to non-Mendelian transmission ratios in their favor. One well-studied example is the *Drosophila* segregation distortion system, where heterozygous SD/SD^{+} males produce almost exclusively SD-bearing progeny. The target of distortion is the *Responder* locus (*Rsp*), which corresponds to an A/T-rich satellite DNA repeat [102]. Strikingly, greater the abundance of satellite repeats, greater is the degree of distortion. A second example is female meiotic drive, where selfish elements that bias random segregation can evolve to exploit the fact that only 1 out of 4 meiotic products becomes a gamete during female meiosis [103, 104, 105]. Centromeric drive is one type of female meiotic drive, whereby expansions of pericentromeric satellite DNAs enhance transmission of linked centromeres to the oocyte, perhaps by capturing a greater fraction of spindle microtubules. Evidence strongly supporting satellite-mediated centromeric drive is observed in hybrids between two *Mimulus* species, where expansion of centromeric repeats is linked to transmission ratio distortion during female meiosis [106].

Genetic conflict provides a very different view of heterochromatin evolution. Unequal crossing over remains the molecular mechanism leading to variation, but if a variant that can selfishly manipulate meiosis or gametogenesis evolves it will rapidly reach high frequency, setting off an arms-race within the host genome to counteract this acceleration.

How does genetic conflict lead to HI? Because genetic conflicts are deleterious to the host species, suppressors will typically arise in the host species to reduce or eliminate the conflict. If this species then hybridizes with a species that has not experienced the conflict, HI can occur if the driver of the conflict segregates away from the suppressor, unleashing the conflict in hybrids. Alternatively, if the suppressor has pleiotropic functions, then divergence of the suppressor may create deleterious interactions in the hybrid and thus cause D-M incompatibility.

Why inferring genetic conflict as a cause of HI is challenging.

(a) Genetic conflicts are transitory. Conflicts that exert strong deleterious effects will be rapidly suppressed or their hosts will go extinct. Traces of past conflicts may then disappear if the driver of the conflict becomes a pseudogene or is purged from the host. For example, transposable elements are often rapidly silenced and accumulate inactivating mutations and deletions [101].

(b) Pleiotropy. Drivers and suppressors of conflicts may be pleiotropic, so that patterns of adaptive evolution in these entities cannot be uniquely attributed to genetic conflict. For example, Presgraves [107] has described a compelling hypothesis that nucleopore proteins have evolved under positive selection because they are recurrently suppressing conflicts involving segregation distorters, with the caveat that the nuclear pore is a highly pleiotropic structure

that may be subject to multiple and distinct evolutionary forces.

(c) Neutral alternatives. Non-coding elements such as heterochromatic repeats have a strong potential to cause genetic conflict, but it is difficult to detect and quantify signatures of selection in non-coding DNA. Furthermore, as we have outlined above, there are wellestablished non-selective mechanisms that are sufficient to account for variation in heterochromatic DNA.

Evidence and examples suggesting genetic conflict as a driver of HI. Perhaps the most compelling evidence is if a hybrid genotype associated with HI also shows evidence of the genetic conflict. The hybrid sterility gene *Ovd* is the best example of this. Crosses between subspecies of *D. pseudoobscura* result in hybrid sons that are almost completely sterile. When these hybrid males are able to reproduce, they produce nearly all daughters. Phadnis and Orr discovered that a single locus, *Ovd*, contributes to both sterility and sex-ratio distortion in hybrid males [84]. While it remains possible that the sex-ratio distortion in hybrids occurs as a pleiotropic consequence of *Ovd* divergence unrelated to suppressing distortion within species, further studies of *Ovd* function in *D. pseudoobscura* can test whether it is involved in suppression of sex-ratio distortion.

There are at least two additional cases of close genetic associations between segregation distorters and hybrid sterility, although the causal genes have yet to be identified. Male mice heterozygous for the *t* haplotype (+/*t*) transmit the *t*-bearing chromosome to nearly all offspring. Four hybrid sterility factors have been mapped to the *t* haplotype region, one of which (*Hst6*) may correspond to the distorter element *Tcd2* locus [108, 109].A second example comes from an introgression from *D. mauritiana* into *D. simulans* which causes both sex-ratio distortion and reduced male fertility. Tao and colleagues have mapped the lo-

cus, named *Tmy*, to a region less than 80 kbp in length [110].

A second line of evidence for conflict is when an HI gene interacts with or co-localizes within the cell with conflict-prone entities. The *OdsH* hybrid male sterility gene provides the strongest example of this second class of HI genes associated with genetic conflict, because as described above the heterochromatic localization properties of OdsH proteins from the hybridizing species have diverged as a consequence of rapid evolution of the DNA binding homeodomain [85]. The implicit assumption linking genetic conflict and *OdsH* is that this divergence in binding site evolved in response to segregation distortion in males or meiotic drive in females. It is unclear though, how *OdsH* might mediate either form of genetic conflict since it is only detected in premeiotic stages of males [85]. Molecular identification of the co-introgressed factors required for hybrid sterility may reveal more about the molecular mechanism underlying the incompatibility, and thereby its evolution.

A reciprocal situation is observed in the incompatibility involving the *Zhr* locus. Here the HI gene is a species-specific heterochromatic repeat, present on the *D. melanogaster* X-chromosome, but absent in *D. simulans* [111]. The paternal *D. melanogaster* X-chromosome, containing the *Zhr* satellite block, fails to segregate in a *D. simulans* maternal background. One hypothesis is that this satellite repeat is a selfish genetic element and the *D. melanogaster* lineage adapted to its expansion by evolving lineage-specific suppressors. The cytoplasm of the *D. simulans* mother would lack these suppressors since it does not contain the *Zhr* satellite block, thereby derepressing this conflict in hybrid daughters. However, as with all satellite evolution the null hypothesis that neutral evolution is responsible for the interspecific difference cannot be excluded. Comparing fitness

differences within *D. melanogaster* of variation in *Zhr* satellite abundance may help to distinguish between these hypotheses.

Going forward with conflict hypotheses. Internal genetic conflict remains a compelling hypothesis to explain the underlying cause of many cases of HI. Alternative hypotheses, however, are not easily refuted. We thus consider approaches for further testing internal genetic conflict hypotheses.

(a) Investigating the intraspecific phenotypes of HI genes. If HI genes are involved in defense against chromosomal drive-based internal genetic conflicts, then one expects to see chromosomal defects during or after meiosis in HI gene mutants. Strikingly, *Hmr*, *Lhr*, *Zhr* and *OdsH* loss-of-function mutant flies form viable stocks, and when investigated more carefully reveal only subtle reductions in female or male fertility [5, 6]. One way to further understand their intraspecific functions will be to test if they have stronger mutant effects in genetic backgrounds that are partially comprised for heterochromatin formation or maintenance. A more direct, but challenging, test of the role of HI genes in suppressing drive is to ask if drive phenotypes are exacerbated in HI gene mutant backgrounds. Such tests will require chromosomes that are polymorphic for defined segments of heterochromatin.

(b) Looking for lineage-specific versus recurrent selection. Is it possible to distinguish between signatures of selection left by genetic conflict compared to responses to ecological or mutational events? Each species has a unique history of invading selfish genetic elements. Each instance of conflict is therefore predicted to be transitory and to quickly evolve suppressors, and thus be confined to single lineages. Lineage-specific divergence thus might be diagnostic of HI genes that have evolved to suppress genetic conflict. For example, all seven non-

synonymous fixed differences of *Ovd* are derived in the *D. pseudoobscura bogota* lineage, consistent with an evolutionary history of segregation distortion being confined to this lineage [84]. In the case of *OdsH*, although the homeodomain has diverged in *D. melanogaster* and *D. simulans*, the acceleration is significantly greater in *D. mauritiana* [8].

In contrast, signatures of selection for *Nup96*, *Nup160*, and *Lhr* are not restricted to only one of the hybridizing lineages [51, 88, 89], and both *Prdm9* and *Hmr* have experienced recurrent selection along multiple lineages [112, 97]. The recurrent evolution of the zinc fingers encoded by *Prdm9* is consistent with adaptation to the recurrent mutational pressure it faces as the determinant of meiotic hotspots but for *Hmr* its evolutionary pattern is impossible to interpret because its molecular function is unknown. A critical question is whether certain genes are recurrently involved in suppressing genetic conflicts. For example, genes involved in nuclear transport have been proposed to be recurrently involved in suppressing segregation distortion [107, 113]. We suggest continuing to test for lineage-specific versus recurrent selection for HI genes, but it will need to be coupled to functional analyses in order to determine its discriminatory power for identifying genetic conflict as a cause of HI.

(c) Linking the signature of conflict to the substitutions causing HI. If genetic conflict is the underlying cause of HI, then the sequence changes in a HI gene that are occurring in response to the conflict should also be causing the HI phenotype. The hybrid lethality gene *Lhr* encodes a heterochromatin protein whose coding sequence has diverged under positive selection between *D. melanogaster* and *D. simulans* [51, 114].We have suggested *Lhr* as another candidate for having co-evolved with selfish heterochromatic elements under genetic

conflict. However, contrary to our expectations, *Lhr* orthologs from both hybridizing species have hybrid lethal activity and foreign-species Lhr proteins localize normally to heterochromatin when expressed in *D. melanogaster* [86]. Further analysis has revealed that despite striking divergence of the protein coding sequence, *cis*-regulatory divergence affecting transcript levels is responsible for the different hybrid lethal properties of *Lhr* orthologs (S.M. and D.A.B., unpublished data).

(d) Quantifying heterochromatin variation in natural populations and its phenotypic consequences. The variation that results in substantial interspecific divergence in heterochromatin must originate within populations, but the nature of this variation is largely unknown. If heterochromatin variation is entirely neutral, then variation should be governed by mutational mechanisms such as unequal crossing over and be continuous. Alternatively, if heterochromatin variation is driven by genetic conflicts, then driving chromosomes should quickly sweep through populations and remove most of the variation. More complicated scenarios may need to be considered, however, because heterochromatic drivers can induce pleiotropic fitness costs that keep them from fixing in populations [106]. Phenotypic assays will therefore need to be coupled to surveys of heterochromatin variation in order to understand fully the population dynamics of heterochromatin.

1.6 Future Directions of HI Studies

1.6.1 Younger and simpler species models.

A chronic concern is whether HI genes were directly involved in causing speciation, or instead evolved after full reproductive isolation. An obvious answer to this problem is to turn to younger species pairs that manifest early-stage HI phenotypes, such as the subspecies of *D. pseudoobscura* we have described above. There are also intriguing cases such as *D. virilis* and *D. americana*, which have relatively high sequence divergence yet seem to have a much simpler genetic basis of hybrid sterility compared to other *Drosophila* [115].

1.6.2 Mechanistic models.

On the other hand, the challenges inherent in rigorously confirming the identity of HI genes and understanding the evolutionary forces and mechanisms underlying HI phenotypes remain formidable even in experimentally tractable organisms. For this reason laboratory model organisms will continue to provide essential discoveries about HI genes, even if their direct relevance to speciation is less clear. The recent discovery of a pair of Caenorhabditis species that produces fertile F1 hybrids opens up great possibilities for molecular genetic approaches to HI in a new experimental model [116].

1.6.3 Genome divergence scans for identifying candidate genes.

Genetic evidence is ultimately required to conclusively identify HI genes, but fine-scale mapping can be slow and laborious. Candidate genes for HI phenotypes like male sterility may be identified by their signatures of rapid evolution, testis expression and possible known roles in male fertility [117].

1.6.4 Intraspecific incompatibilities.

The neo-Darwinian synthesis posits that intras-pecific variation and interspecific divergence exist on a continuum. It is less clear if this also holds for incompatibilities, because alleles causing intraspecific incompatibility will be strongly selected against unless they exist in isolated populations. However, a number of intriguing cases of intraspecific incompatibilities have been reported [118, 119, 120, 121]. Yeast again has provided exceptional opportunities for unraveling the genetic basis of incompatibility, with an inter-strain defect in DNA mismatch repair being shown to result from a single amino-acid change in each of two interacting proteins [120]. A substantial number of sterile and lethal interactions have recently been discovered between *D. melanogaster* populations [122]. Assessing the relevance of such intraspecific incompatibility to interspecific HI then requires an understanding of the patterns of gene flow between populations.

1.6.5 Embracing complexity.

Many HI phenotypes involve multiple genes that display complex epistatic interactions. Untangling these phenotypes is not just necessary in order to identify all the causal HI genes. Rather, important general properties such as the degree of dominance cannot be accurately assessed without considering the full set of epistatic interactions involved [123]. The challenge of fully understanding genetic complexity of HI phenotypes will be resolved in part by the ability to genotype more recombinant progeny and phenotype them with more sophisticated assays in QTL mapping experiments.

CHAPTER 2

RECURRENT POSITIVE SELECTION OF THE DROSOPHILA HYBRID INCOMPATIBILITY GENE *HMR*[1]

2.1 Introduction

Progress has been made in understanding the genetics of speciation by reducing the complexities of speciation to investigation of the genetic basis of reproductive isolation [124]. A population undergoing divergent evolution can ultimately result in the creation of reproductively isolated populations, or new species. The evolution of reproductive isolation requires the establishment of barriers to gene flow, often multiple barriers acting together at various stages in the life cycle of the organism. Hybrid incompatibility (HI), the sterility and inviability of interspecific offspring, is a postzygotic barrier to gene flow. As it is relatively easy to measure the viability and fertility of hybrid progeny, HI has been more amenable to genetic dissection than other reproductive isolating mechanisms. The study of HI loci is also of great interest because it addresses how developmental pathways may diverge between taxa, a process that characterizes both genome evolution and speciation.

Five HI genes have been described: *Xmrk-2* from the fish *Xiphophorus*, and the Drosophila genes *OdsH*, *Hmr*, *Nup96* and *Lhr*, [125, 8, 126, 88, 51]. Based on sequence homology two of these genes might be involved with transcriptional regulation or chromatin binding (*OdsH* and *Hmr*), *Nup96* encodes a component

[1]This is a manuscript published as Maheshwari S., Wang J. & Barbash D.A. (2008) *Mol. Biol. Evol.* **25**(11):24212430. Author contributions are as follows: S.M. J.W. & D.A.B conceived and designed the experiments. S.M & D.A.B performed the population sequencing. S.M. analyzed data for all the tables and Fig. 2.1 and J.W. analyzed data for Figs. 2.2 & 2.3. S.M. & D.A.B. wrote the paper.

of the nuclear pore complex, *Lethal hybrid rescue* (*Lhr*) encodes a heterochromatin-associated factor, and *Xmrk-2* encodes a receptor tyrosine kinase. Mitochondrial cytochrome *c* oxidases show significant reductions in activity in combination with nuclear-encoded cytochrome *c* proteins that derive from different populations of copepods, making their respective genes strong candidates for causing reduced fitness in copepod hybrids [71]. There is, therefore, no single functional class of genes causing hybrid incompatibility. However, four of these HI genes (*OdsH*, *Hmr*, *Nup96* and *Lhr*) have diverged rapidly under positive selection [8, 88, 127, 51]. The biological significance of HI loci being the target of adaptive evolution is unclear, because if HI evolves as a secondary byproduct of intraspecific evolution then the phenotype being selected for is unlikely to be hybrid incompatibility. One possible explanation is that if mutations causing HI are rare, then HIs will tend to occur in genes undergoing high rates of substitution. Positive selection would be the engine driving a high substitution rate. Positive selection also implies that the genes may be changing in function in a way that causes developmental breakdown in hybrids. These findings raise several intriguing questions. Is the selection pressure on HI genes limited only to the hybridizing species, or have HI genes experienced recurrent adaptive evolution in other species? Have HI genes changed in their structural properties as well as in primary sequence? Does analysis of HI genes resolve otherwise ambiguous phylogenetic relationships?

Here we address these questions for *Hybrid male rescue* (*Hmr*), an HI gene identified in *Drosophila melanogaster*. Matings between *D. melanogaster* mothers to fathers from its sibling species *D. mauritiana*, *D. simulans* and *D. sechellia* produce the same HI phenotype: semiviable but sterile daughters and lethal sons [128, 129].*Hmr* was identified by a loss-of-function mutation in *D. melanogaster*

(*Hmr*[1]) that rescues F1 hybrid sons from each of these interspecific crosses [130, 7]. Population genetic analysis revealed that *Hmr* has diverged under positive selection in both *D. melanogaster* and *D. simulans* [127]. In order to obtain a more comprehensive view of *Hmr* evolution in this study we have: 1) analyzed Myb/SANT-like domain in ADF1 (MADF) domains of *Hmr* orthologs from 14 species within the *Drosophila* genus in order to detect possible changes in DNA or chromatin binding; 2) applied maximum likelihood phylogenetic analysis on 7 species within the *melanogaster* subgroup and 3) generated and analyzed population sampling data from the three sibling species and from the *melanogaster* subgroup species pair *D. yakuba* and *D. santomea*.

2.2 Materials and Methods

2.2.1 Identification and alignment of orthologs.

Hmr orthologs in *D. simulans*, *D. mauritiana*, *D. sechellia* and *Drosophila erecta* were described previously [7, 127]. Additional *Hmr* orthologs were identified here using *D. melanogaster* HMR in TblastN searches of the trace archives or of preliminary assembled contigs from the various Drosophila genome projects [131]. All putative orthologs were reciprocally blasted back to *D. melanogaster* and *Hmr* was identified as the highest scoring hit. We also looked for conservation of synteny using the flanking genes *CG2124* and *Rab9D* (*CG32678*). Synteny was conserved in *D. yakuba*, *D. pseudoobscura* and *D. willistoni*. Only a *Rab* homolog was found adjacent to *Drosophila ananassae* and *Drosophila virilis Hmr*. Neither gene was found adjacent to *Hmr* in *Drosophila mojavensis*. Our designa-

tion of *Hmr* orthologs matches those of the published genome assemblies [131].

Partial sequence of *Drosophila persimilis Hmr* was obtained from *National Center for Biotechnology Information* (NCBI) trace archives using discontinuous MegaBLAST and then completed by sequencing PCR products from DNA extracted from a single *D. persimilis* male from the WSH3 strain that was used for the whole-genome shotgun sequencing project. *Drosophila teissieri Hmr* was sequenced from template DNA generously provided by Dr. John Pool (Cornell University) (GenBank accession number FJ151263). The primers used were the most robust ones from the *D. yakuba*/*D. santomea* population study as well as *D. teissieri*-specific primers.

The gene structures of *Hmr* orthologs were predicted by GeneWise software [132] guided by *D. melanogaster Hmr* and manually checked for exon-intron conservation. We were unable to identify a homologous exon 1 in *D. pseudoobscura*, *D. ananassae*, *D. virilis* or *D. mojavensis*. We therefore annotated *Hmr* in these species as having the longest conceptual open reading frame initiating in the large exon that is orthologous to exon 2 of *D. melanogaster*. Some of our *Hmr* annotations differ from Clark et al. (2007), but these differences do not affect the MADF domains analyzed in Table 2.1.

2.2.2 Population samples.

Hmr was isolated from twelve lines of *D. simulans* including 5 lines used in a previous study [127] and an additional 7 lines collected in Zimbabwe, Africa. Twelve lines of *D. mauritiana* were obtained from Dr. Shun-Chern Tsaur (Academia Sinica, Taipei, Taiwan; "W" lines) or from the Tucson Drosophila Stock Center

(other lines). Five lines of *D. sechellia* were obtained from the Tucson Drosophila Stock Center. Primers for PCR amplification and sequencing are described in [127]. GenBank accession numbers for these sequences are *D. simulans* FJ151256 - FJ151262, *D. mauritiana* FJ151229 - FJ151240 and *D. sechellia* FJ151252 - FJ151255. *Hmr* was also isolated from 11 *D. yakuba* and 11 *D. santomea* strains. Flies for the different strains of *D. yakuba* and *D. santomea* were obtained via Dr. David Begun (Univ. of California, Davis) from collections of Dr. Peter Andolfatto (Univ. of California, San Diego) and Dr. Manyuan Long (Univ. of Chicago). *D. yakuba* lines were collected in Cameroon and are described in [133]; the *D. santomea* samples were from several populations. PCR primers were designed from the genome sequence of *D. yakuba*, to cover the entire *Hmr* gene in five overlapping amplicons of approximately 1.2 kb each. Multiple attempts using several primer sets to PCR amplify the 3'-most amplicon failed from many of the samples. This necessitated that we restrict our population genetic analysis to the *Hmr* region ending 1067 bp upstream of the stop codon in the reference *D. yakuba Hmr* CDS. GenBank accession numbers for these sequences are *D. yakuba* FJ151264 - FJ151274 and *D. santomea* FJ151241 - FJ151251. A low-complexity region that contained polymorphic indels in both species was excluded from our analyses, it corresponds to amino-acid positions 802 to 885 and 787 to 855 for *D. yakuba* and *D. santomea* reference sequences, respectively. Due to the difficulty in sequencing the 3'-most amplicon from single-fly preps, a complete *Hmr* gene sequence from *D. santomea* was generated by synthesizing a composite allele. The 3'-most block was amplified and sequenced from DNA extracted from 25 *D. santomea* flies (GenBank accession number FJ151275). This was joined to one *D. santomea Hmr* allele chosen at random.

DNA from a single male fly for each strain was used as a template in the

PCR reactions in order to obtain a single allele of the X-linked *Hmr*. DNA was prepared by the method of SDS lysis followed by phenol-chloroform extraction. The PCR products were purified either with Qiagen PCR cleanup columns or gel purified using the QIAEX II Gel Extraction Kit from Qiagen. The PCR products were then sequenced directly using Big Dye version 1.1, 3.1 (Applied Biosystems) reagents on an ABI capillary sequencer.

Sequences were aligned using MegAlign from the Lasergene v.6 package (DNASTAR, Inc. Madison, WI) and the alignments corrected by eye.

2.2.3 Analyses of divergence.

ClustalW was used for multiple sequence alignment of *Hmr* within and between species [134]. All the coding sequence alignments were obtained by first aligning their protein products. Using RepeatMasker [135], two low-complexity regions located in the 2nd and 4th exons were found in *melanogaster* subgroup species and were removed from some analyses to ensure accurate alignment. The total length of removed material ranged from 22 amino acids in *D. melanogaster* to 122 amino acids in *D. yakuba*. Two low-complexity regions were also found in the first exon of *D. persimilis* (81 amino acids) and *D. pseudoobscura* (121 amino acids). Alignments without these low-complexity regions were used for the construction of the eight-species phylogeny, and the PAML analyses (Figure 2.3). The complete *Hmr* sequences from *D. melanogaster*, *D. simulans*, *D. mauritiana* and *D. sechellia* were used for McDonald-Kreitman (MK) tests (Table 2.3). For *D. yakuba* and *D. santomea* alignments without the low-complexity regions were used for MK tests (Table 2.3).

Neutrality tests were carried out in DnaSP v.4.5 [136]. Tests of whether synonymous sites are evolving toward preferred or unpreferred codons were made using the method of DuMont et al. (2004)[137] with the "Biased" mutations options. Significance was tested using Fisher's Exact Test (two-tailed).

2.2.4 Phylogenetic analysis.

Phylogenetic trees were built by MEGA 3.1 using parsimony and Neighbor-joining methods [138]. PAML was used for the maximum-likelihood method of phylogenetic analysis [139]. The lineage-specific models in PAML allow for the variation of D_N/D_S ratios among different lineages. The M0 (one-ratio) model was compared with a two-ratio model as well as a free-ratio model along each lineage. P values were calculated in R 2.2.0 using the likelihood ratio test of each comparison [140]. Figures of phylogenetic trees were prepared by retracing the primary images in Adobe Illustrator.

2.2.5 Structural analyses.

Secondary structure predictions were made using Jpred [141]. The charge and isoelectric points (pI) were predicted using the Editseq program which is part of the Lasergene v.6 package (DNASTAR, Inc.).

2.3 Results

2.3.1 Domains and structure of *Hmr*.

Hmr from *D. melanogaster* encodes a predicted protein of 1,413 amino acids, in which two MADF domains were identified previously [7]. The MADF domain (MYB-related ADF domain) was discovered in the Drosophila protein ADF1 based on its sequence similarity to the DNA binding domain of MYB, and is required for the DNA-binding activity of ADF1 [142, 143]. The predicted MADF secondary structure is also similar to the SANT domain, which is found in a large number of DNA and chromatin-associated proteins [144]. We identified orthologs of *Hmr* from 14 *Drosophila* species using previously published work [127], our sequencing here, and preliminary assemblies of whole-genome shotgun data [131]. While examining these *Hmr* orthologs we identified two additional candidate MADF domains (Figure 2.1A). These four HMR MADF domains are generally highly conserved among all *Drosophila* species.

A close comparison of the *D. melanogaster* HMR MADF domains to each other and to the ADF1 MADF domain reveals that the third and fourth putative MADF domains contain insertions and differ at certain conserved residues, which raises the question of whether these four domains have the same function. Interestingly, we find considerable variation in the predicted charge and isoelectric point of each domain (Figure 2.1B). MADF1 is highly positively charged and thus most closely resembles DNA-binding domains found in MYB or ADF1, while MADF3 is negatively charged and thus is more similar to a chromatin-binding domain such as the SANT domain from ISWI. MADF2 and MADF4 have a significantly lower charge compared to the canonical ADF1 MADF do-

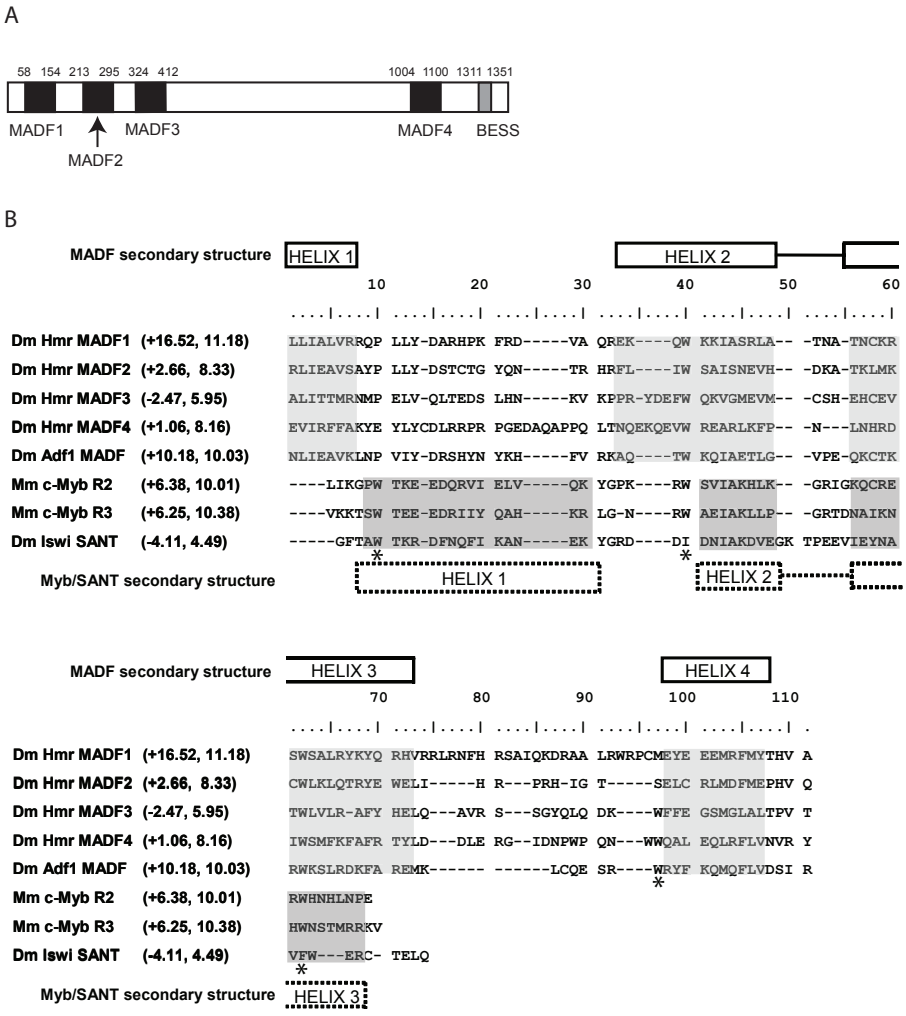

Figure 2.1: Domain structure and MADF domains of Hmr. A) The positions of four MADF domains and the putative BESS domain [51] encoded by *D. melanogaster Hmr* are shown. B) Alignment and comparison of MADF, MYB and SANT domains. The four MADF domains from *D. melanogaster* HMR and the MADF domain from *D. melanogaster* (Dm) ADF1 were aligned with the second and third MYB domains from *M. musculus* (Mm) c-MYB, and the SANT domain from *D. melanogaster* ISWI. This alignment was manually adjusted to the published alignment of MYB and SANT domains [145, 146]. Asterisks underneath the alignment indicate the conserved tryptophan residues, and the shaded boxes indicate predicted α-helical regions. The ionic properties of each domain are indicated within parentheses after the domain name as (charge, isoelectric point), respectively. HMR MADF1 and ADF1 MADF have ionic properties similar to the DNA-binding MYB domain, while HMR MADF3 is distinct in being more similar to the chromatin-associated SANT domain. For comparison, DIP3, another MADF-domain-containing transcription factor that has been experimentally shown to bind DNA [145], has a MADF domain with a charge of +8.25 and an isoelectric point of 10.52, which is again similar to the MYB domain. The NCBI Entrez GeneIDs are: Mm c-Myb, 17863; Dm Iswi, 36390; Dm Adf1, 47082; Dm Dip3, 53579.

main, although their pI values are still consistent with a putative DNA binding function. These results imply that each of the four HMR MADF domains may have unique functions with respect to DNA or histone association. These differences also suggest that HMR may bind to both DNA and the protein component of chromatin.

To determine if these unusual MADF domains are unique to *D. melanogaster* HMR, we analyzed the MADF domains encoded by *Hmr* orthologs from 13 other species in the *Drosophila* genus (Table 2.1). A clear trend was observed, with the first domain within each species' HMR resembling a canonical MADF domain consistent with DNA-binding function. The remaining three MADF domains, however, have variable ionic properties. *Hmr* orthologs from taxa within the *melanogaster* group are in general similar to the *D. melanogaster* ortholog, with only the third MADF domain having a net negative charge, suggestive of chromatin binding. One exception is the fourth predicted MADF domain from *D. erecta* that also has a net negative charge. *Hmr* orthologs from other *Sophophora* species (*D. pseudoobscura*, *D. persimilis* and *D. willistoni*) do not encode a negatively charged third MADF domain. The most divergent ortholog is from *D. mojavensis* with three MADF domains having a net negative charge.

We conducted a similar analysis of MADF domains identified by the SMART database for other *D. melanogaster* genes, in order to determine how common are MADF domains with a predicted net negative charge. We only found two other genes which encode MADF domains with isoelectric points below seven: *CG31627* (pI = 6.75) and *CG1603* (two MADF domains with pIs = 6.75 and 9.3). We conclude that most MADF domains are likely to be involved in DNA binding, while *Hmr* encodes unusual MADF domains with potential chromatin

Table 2.1: Predicted charge and isoelectric points (pI) of MADF domains within *Hmr* orthologs from 14 *Drosophila* species.

	MADF1		MADF2		MADF3		MADF4	
	Charge	pI	Charge	pI	Charge	pI	Charge	pI
D. melanogaster	16.52	11.18	2.66	8.33	**-2.47**	**5.95**	1.06	8.16
D. simulans	11.85	10.55	2.52	8.51	**-0.61**	**6.6**	2.59	9.21
D. mauritiana	12.02	10.32	2.72	8.78	**-0.61**	**6.6**	2.56	8.82
D. sechellia	11.18	10.37	2.69	8.51	**-0.61**	**6.6**	2.56	8.83
D. yakuba	13.19	11.55	3.55	9.24	**-2.5**	**5.95**	1.76	8.52
D. santomea	13.19	11.64	3.55	9.24	**-1.51**	**6.32**	1.76	8.52
D. teissieri	12.02	11.54	4.89	9.74	**-3.34**	**5.85**	2.39	9.16
D. erecta	12.36	10.35	4.52	9.39	**-2.44**	**5.96**	**-0.57**	**6.62**
D. ananassae	9.86	10.31	2.55	8.72	**-2.67**	**5.58**	0.76	7.54
D. pseudoobscura	12.36	10.35	2.25	9.13	1.72	8.19	2.39	8.8
D. persimilis	9.36	10.03	1.42	8.48	1.55	8.17	0.43	7.42
D. willistoni	8.88	10.31	2.25	9.1	2.12	8.17	1.43	8.48
D. mojavensis	16.69	11.41	**-1.24**	**6.49**	**-0.17**	**6.5**	**-0.4**	**6.77**
D. virilis	16.52	11.56	4.74	9.62	**-1.87**	**5.69**	3.09	9.18

MADF domains with predicted negative charges and acidic isoelectric points are in bold.

binding properties.

Nearly all *Hmr* orthologs also encode simple amino-acid repeats, consisting predominantly of serine, alanine and proline. Such simple-sequence repeats are over represented among transcription factors [147, 148]. Sequencing of multiple *Hmr* alleles from the *D. yakuba* and *D. santomea* species pair (see below) revealed a unique microsatellite-like polymorphism within the coding region. The kernel of this repeat is present within the *melanogaster* subgroup, but the expansion is restricted to the lineage leading to *D. yakuba*, *D. santomea* and *D. teissieri*. The expansion has resulted in a tandem repeat consisting of nearly perfect alternating "SAT" and "QAA" residues, ranging from 63 amino acids to 87 amino acids in length.

2.3.2 *Hmr* phylogenetic pattern in *melanogaster* subgroup.

Outside of the *melanogaster* subgroup the predicted HMR protein is highly diverged and thus impossible to align fully. Therefore we aligned *Hmr* only from 8 species within the *melanogaster* subgroup. Phylogenetic trees were built using both the Neighbor-joining and maximum-parsimony methods. These methods produced similar results, except for the grouping of the *D. melanogaster* sibling species *D. simulans*, *D. mauritiana* and *D. sechellia*. We therefore obtained a population data set from these 3 species to further explore their phylogenetic relationship (Table 2.2).

Phylogenetic reconstruction of the population data set using maximum parsimony provided support for *D. sechellia* branching off prior to the split of *D.*

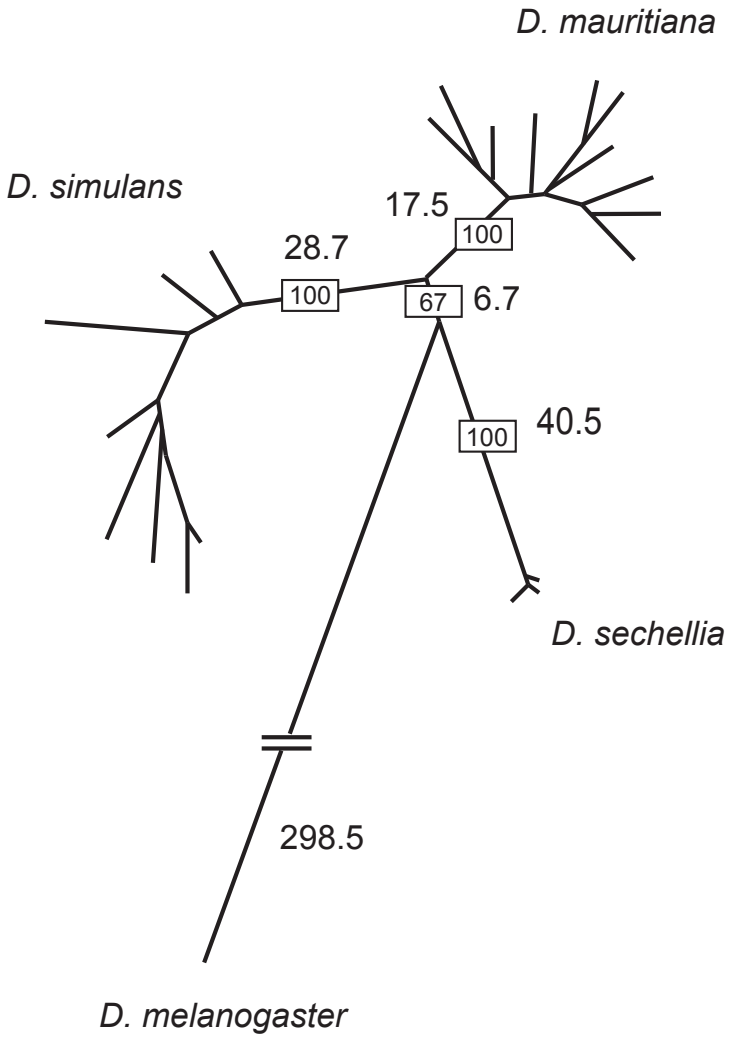

Figure 2.2: A maximum-parsimony phylogenetic tree constructed using *Hmr* population samples from the three *D. melanogaster* sibling species. Bootstrap values of major branches are shown in the rectangles (100 replicates). The lengths of major branches, which are proportional to the number of nucleotide changes for the gene, are indicated along the branches. *Hmr* has clearly diverged and shares no common alleles among the three sibling species.

simulans and *D. mauritiana*, albeit with relatively low bootstrap support (Figure 2.2). We therefore further analyzed all the sites that have an unambiguous phylogenetic signal. The definition of unambiguous sites follows Ting et al. (2000) [149]: a site is defined as unambiguous only when two of the three species share a derived nucleotide with none of their alleles having the ancestral nucleotide, while the third species has the ancestral nucleotide with no alleles having the derived one. We found a total of 10 unambiguous sites for *Hmr*. Among them, 6 sites support the grouping of *D. simulans* and *D. mauritiana*, 3 sites support the grouping of *D. simulans* and *D. sechellia*, and only one site supports the grouping of *D. mauritiana* and *D. sechellia*. This result is consistent with both of our maximum-parsimony phylogenetic trees (Figure 2.2).

2.3.3 Tests for selection: Phylogenetic analysis among multiple *Drosophila* species.

Using a single outgroup sequence *Hmr* was previously inferred to have an increased ratio of non-synonymous to synonymous substitutions (D_N/D_S ratio) along the branches leading to *D. melanogaster* and its sibling species [127]. In order to obtain a more comprehensive view of *Hmr* evolution in the *melanogaster* subgroup we applied a maximum likelihood analysis on 7 species (Figure 2.3). Because we obtained conflicting phylogenies for *D. simulans* and its sister species *D. mauritiana* and *D. sechellia* from two different phylogenetic methods, we excluded *D. sechellia* in this PAML analysis to avoid possible artifacts caused by using the incorrect evolutionary history. Using a free-ratio model we confirmed that the estimated D_N/D_S ratios for *Hmr* have generally increased along branches

leading to *D. melanogaster* and its sibling species, relative to other lineages in the subgroup. We found D_N/D_S ratios of approximately one or higher along the lineages leading to *D. melanogaster*, *D. simulans* and *D. mauritiana*. These values are similar but not identical to those of Barbash et al. (2004), due to the inclusion of different species sequences in the two studies. Note also that the values for D_N and D_S for *D. mauritiana* were erroneously switched in Figure 3 of Barbash et al. (2004).

D_N/D_S ratios were also relatively high for other lineages of the subgroup, with the striking exception of the *D. yakuba* and *D. santomea* lineages after the split from their common ancestor, where D_N/D_S was approximately 0.2.

2.3.4 Tests for selection: Evidence for positive selection in *D. simulans* and *D. mauritiana*.

We analyzed population genetic datasets from different species pairs and groups in order to further explore the very different estimations of divergence in Figure 2.3. We first examined the sibling species of *D. melanogaster* (Table 2.2). Polymorphism values, including the much lower level for *D. sechellia*, were generally consistent with observations from other genes [150].

The MK test [151] was carried out for *D. simulans* and *D. mauritiana* and rejects the null hypothesis of neutral evolution with a highly significant P-value (Table 2.3). This comparison shows a particularly high amount of nonsynonymous substitutions relative to synonymous substitutions. We then polarized substitutions using *D. melanogaster* as an outgroup and rejected neutrality along

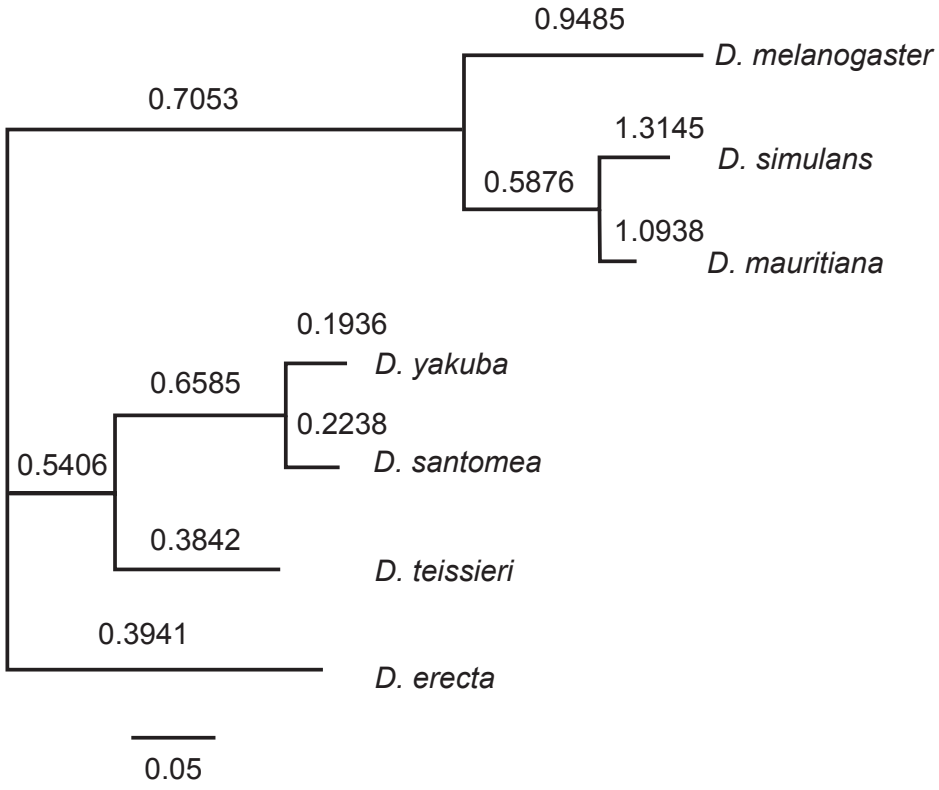

Figure 2.3: A maximum-likelihood phylogenetic tree of *Hmr* built by the free-ratio model in PAML. The likelihood of this model was significantly better than the one-ratio model (2Δ1= 54.224, df=11, p¡10-7). The tree length is defined as the number of nucleotide substitutions per codon. The number shown above each lineage is the estimated D_N/D_S value of that lineage.

Table 2.2: Summary of *Hmr* polymorphism data.

	D. simulans	*D. mauritiana*	*D. sechellia*	*D. yakuba*	*D. santomea*
No. of alleles analyzed	13	12	5	11	11
Nucleotide diversity (Pi)	0.00878	0.00606	0.00117	0.00815	0.0028
Total variable sites	133	108	12	126	38
Variable sites within coding sequence	110	90	10	104	30
Total number of singletons	35	55	7	82	21
Number of synonymous singletons	15	30	2	45	10
Number of nonsynonymous singletons	15	17	3	20	6

both the *D. simulans* and *D. mauritiana* lineages. The significance of these tests are most likely caused by an excess of nonsynonymous substitutions. We tested an alternative hypothesis that departures from neutrality may be due to selection on synonymous sites for preferred codons [137]. Tests were not significant for either *D. mauritiana* (p=0.339) or *D. simulans* (p=0.115) using *D. melanogaster* as an outgroup.

Pairwise comparisons of both *D. simulans* and *D. mauritiana* with *D. sechellia* also reject neutral evolution (Table 2.3). These tests are not independent from the above *D. simulans*-*D. mauritiana* comparisons, but rather reinforce the inference of positive selection on those two lineages. There is little power to test for non-neutral evolution along the *D. sechellia* lineage due to the very low amount of polymorphism in this species (Table 2.2), and not surprisingly, MK tests did not reject the null hypothesis for the *D. sechellia* lineage (data not shown).

In combination with the D_N/D_S estimates in figure 2.3 we conclude that *Hmr* has continued to diverge under positive selection along both the *D. mauritiana* and *D. simulans* lineages after the divergence of their common ancestor from *D. melanogaster*.

2.3.5 Tests for selection: Evidence for positive selection in the *D. santomea* lineage.

The MK test for *D. yakuba* and *D. santomea* also rejected the null hypothesis of neutral evolution. Like in the aforementioned tests with *D. simulans* and *D. mauritiana* there is a higher relative ratio of nonsynonymous to synonymous

Table 2.3: McDonald-Kreitman tests of neutrality.

		Divergence		Polymorphism		
Pairwise Comp.	Species pair or lineage	Syn	Nonsyn	Syn	Nonsyn	Fisher's exact test P
1	*D. simulans vs D. mauritiana*	3	28	90	105	6.138E-05***
	D. simulans lineage	2	16	48	61	8.537E-03**
	D. mauritiana lineage	1	8	45	44	0.03366*
2	*D. simulans vs D. sechellia*	16	43	53	68	0.03450*
3	*D. mauritiana vs. D. sechellia*	16	42	49	49	0.007209**
4	*D. yakuba vs. D. santomea*	22	22	92	41	0.02889*
	D. yakuba lineage	10	7	72	32	0.4113
	D. santomea lineage	10	13	21	8	0.04823*
5	*D. yakuba vs. D. teissieri*	93	145	72	32	3.301E-07***
	D. yakuba lineage	40	70	71	32	2.597E-06***
6	*D. santomea vs. D. teissieri*	93	146	21	8	1.113E-03**
	D. santomea lineage	43	73	22	8	4.382E-04***

Six pairwise comparison were tested, using total polymorphism in comparisons 1-4. Within some pairwise comparisons both lineages (comparisons 1 and 4) or one lineage (comparisons 5 and 6) were then tested by polarizing each lineage using an outgroup sequence. For *D. simulans* and *D. mauritiana* (comparison 1), *D. melanogaster Hmr* was used as the outgroup sequence; for *D. yakuba* and *D. santomea* (comparison 4), *D. teissieri Hmr* was used as the outgroup sequence; for comparisons 5 and 6, *D. erecta Hmr* was used as the outgroup sequence. Syn = Synonymous sites ; Nonsyn = Nonsynonymous sites

variation between species relative to within species. Polarization of these data relative to *D. teissieri* revealed non-neutral evolution exclusively on the lineage leading to *D. santomea*. Analysis of the synonymous substitutions suggested that there is not an excess of substitutions leading to preferred codons for either species, using *D. teissieri* as an outgroup [p=0.286 for *D. yakuba*; p=1.000 for *D. santomea*].

2.3.6 Tests for selection: Evidence for positive selection in the lineage leading to *D. yakuba* and *D. santomea*.

While estimated D_N/D_S ratios were low along the branches leading to both *D. yakuba* and *D. santomea*, D_N/D_S had a high value of 0.6585 in the lineage leading to their common ancestor (Figure 2.3). We therefore extended our MK test analyses to a pairwise comparison of each species with *D. teissieri*. In both cases we rejected neutral evolution. These results were not due to selection on synonymous sites for preferred substitutions (p = 0.845 for *D. yakuba* using *D. erecta* as the outgroup; p = 0.839 for *D. santomea* using *D. erecta* as the outgroup; similar results were obtained when using *D. melanogaster* as the outgroup [data not shown]). We also polarized these MK test results, and again rejected neutral evolution for both the *D. yakuba* and *D. santomea* lineages. These tests of the *D. yakuba* and *D. santomea* lineages are clearly not independent, as the majority of the substitutions occurred before the speciation of *D. yakuba* and *D. santomea*. However, our results do strongly suggest that *Hmr* diverged under positive selection in the common ancestor of these two species.

2.3.7 *Hmr* in other species.

In order to obtain a view of *Hmr* evolution outside of the *melanogaster* subgroup we assembled and analyzed *Hmr* orthologs from the species pair *D. persimilis* and *D. pseudoobscura*. We estimated D_N/D_S between these species to be 1.174. This high value is suggestive of possible adaptive evolution between these species but will require further analysis, in part because of the low level of divergence between these species (DS=0.0345).

2.4 Discussion.

2.4.1 Recurrent positive selection of *Hmr*.

Hmr causes lethality in hybrid progeny of *D. melanogaster* females mated to males of its sibling species. This lethality reflects a divergence in function of *Hmr* between these species because lethality is only caused by Hmr^+ from *D. melanogaster* and not by Hmr^+ from *D. simulans* or *D. mauritiana* [127]. These genetic observations might suggest that *D. melanogaster* Hmr^+ has diverged from its ancestral state but sibling-species Hmr^+ has not. In contrast to this simple evolutionary scenario, *Hmr* has diverged extensively along both the *D. melanogaster* and *D. simulans* lineages, and has done so in a manner consistent with positive selection rather than neutral evolution [127].

Here we have found that *Hmr* has continued to diverge under positive selection between the sibling species *D. simulans* and *D. mauritiana*. Both lineages have high D_N/D_S values (Figure 2.3), and polymorphism samples show an ab-

sence of allele sharing (Figure 2.2) and a rejection of neutral evolution by MK tests (Table 2.3). Our data also demonstrate positive selection along the lineage leading to the common ancestor of *D. yakuba* and *D. santomea*. This branch has a high D_N/D_S value (Figure 2.3), and pairwise tests of both species with *D. teissieri* clearly reject neutral evolution (Table 2.3). We suggest that these cases demonstrate that *Hmr* has experienced independent episodes of recurrent adaptive evolution in the *melanogaster* subgroup along at least three evolutionary branches: between *D. simulans* and *D. mauritiana*, between *D. melanogaster* and the common ancestor of its sibling species (the ancestor of the *simulans* clade), and between *D. teissieri* and the common ancestor of *D. yakuba* and *D. santomea*. Our analysis of the subsequent divergence of *D. yakuba* and *D. santomea* also suggests that nonneutral evolution continued in the *D. santomea* lineage. D_N/D_S values are low on both branches, but the MK test between these species rejects neutral evolution and polarization confines this signal to the *D. santomea* branch (Table 2.3). Might *Hmr* cause hybrid incompatibility between any of these species pairs? Introgression studies in *D. mauritiana*/*D. simulans* hybrids [28] and QTL analysis in *D. yakuba*/*D. santomea* hybrids [37] have found evidence for genes contributing to male sterility in or near the respective regions corresponding to *D. melanogaster* cytological region 9D, where *Hmr* is located. Whether *Hmr* contributes to these phenotypes remains speculative, since the mapping resolution was relatively low, and these and other studies [26] suggest that there is a high density of X-linked hybrid male sterility factors in *Drosophila*.

Our results raise the question of how common is recurrent adaptive evolution, for orthologs of other HI genes, and more generally for other classes of genes. The gene *Odysseus* (*OdsH*) causes male sterility in *D. simulans*/*D. mauritiana* hybrids and has a large excess of nonsynonymous substitutions com-

pared to synonymous substitutions, strongly suggesting that it diverged between these species under positive selection [8]. In contrast, *OdsH* orthologs from species of the *Drosophila montium* subgroup have low D_N/D_S values, suggesting that it is evolving under purifying selection in these species [152]. A similar pattern was seen for the Drosophila innate immunity gene *Relish*, which shows evidence for adaptive evolution in *D. simulans* but not in several other *Drosophila* species pairs, based on population genetic sampling [153, 154]. In contrast, population genetic analyses indicate that the spermatogenesis gene *roughex* has undergone at least two independent rounds of recurrent adaptive evolution, between *D. melanogaster* and *D. simulans*, and between *D. yakuba* and *D. santomea* [155].

2.4.2 Phylogeny of the *simulans* species complex and positive selection.

The phylogeny of the *simulans* complex is not well-resolved, because for many genes multiple alleles from different species often group with each other instead of resolving only within their species (Kliman et al. 2000). In contrast, phylogenetic analysis of our *Hmr* population data set fully resolves *D. simulans*, *D. mauritiana* and *D. sechellia* (Figure 2.2). Strongly supported phylogenies were previously obtained for *OdsH* [149] and the retroviral envelope-derived gene *Iris* [156]. However the results differed, with *OdsH* supporting the same pattern as *Hmr* with *D. sechellia* branching off first, whereas *Iris* supports a different phylogeny where *D. mauritiana* branches off first.

Two explanations may help explain these discrepancies. The first is the hy-

pothesis that reproductive isolation does not occur as a single event but rather that different regions of the genome will become isolated at different times during nascent speciation [157]. In this view, the genomes of well-defined species will be a mosaic of regions that have different histories of isolation. While both *Hmr* and *OdsH* are X-linked, they are not particularly close to each other (cytological regions 9D and 16D, respectively). Considering that a region less than 2 kilobases away from *OdsH* showed a distinct phylogenetic pattern [149], *Hmr* and *OdsH* cannot share the same phylogenetic pattern due to linkage. There are also no known large inversions between the *simulans* complex species. Therefore under the mosaic genome hypothesis this similarity between *Hmr* and *OdsH* in their phylogenetic pattern would likely be coincidental.

A second possible explanation for the phylogenetic discrepancies is that these genes may have experienced selection in different subsets of the three simulans complex species [156]. It is difficult to directly test this possibility for *OdsH*, *Hmr* and *Iris* because of the use of different methods to detect selection and different data sets. *OdsH* shows strong evidence for positive selection between *D. simulans* and *D. mauritiana* based on a high D_N/D_S ratio within its homeodomain but the *D. sechellia* lineage has not been examined. *Iris* shows evidence of positive selection between *D. melanogaster* and the common ancestor of the *simulans* complex, and codon-based models rejected neutral evolution among 12 *Drosophila* species, but selection specifically between *simulans* complex species has not been detected.

2.4.3 Assessment of MADF domain functional properties.

The DNA-binding function of the MYB domain has been biochemically well-established in a diverse set of transcription factors from a wide variety of eukaryotes [158, 159]. The SANT domain was discovered as a conserved region, closely related to the MYB domain, in the chromatin remodelers and/or transcriptional co-factors SWI3, ADA2, N-CoR and TFIIIB [144]. Where the MYB domain presents a basic surface that contacts the negatively charged DNA phosphate backbone, the SANT domain presents a distinctly acidic surface that is much more likely to contact positively charged histone tails. This difference is also evident in the general ionic properties of each domain, such as c-MYB R2 (+6.38, pI 10.01) compared to ISWI SANT (-4.41, pI 4.49) [160] (Figure 2.1B).

MADF domains from two *Drosophila* proteins, ADF1 and DIP3, have been shown to bind directly to DNA [143, 145], and both are positively charged (Figure 2.1B). MADF domains from most other Drosophila proteins are also positively charged, as are 3 out of 4 of the predicted MADF domains from HMR in most species (Table 2.1). HMR MADF3, however, is negatively charged in many species, including *D. melanogaster*, suggesting that HMR may have both DNA and chromatin binding properties.

2.4.4 Molecular signatures of an HI gene.

Analyses of a handful of experimental model organisms have led to the identification of many conserved genes that have critical structural and regulatory roles in species throughout the eukaryotic kingdom. Will the identification of speciation genes in model organisms such as Drosophila be similarly generalizable?

This remains an open question because our understanding of the speciation process is extensive but the collection of known speciation genes is sparse.

Most HI genes identified to date show evidence of positive selection [8, 88, 127, 51]. These observations therefore suggest that candidate HI genes may be identifiable from whole-genome comparisons by the criteria of high D_N/D_S ratios and non-neutral evolution.

The high evolutionary rate of *Hmr* may also have led to it having unique functional properties. Three out of the four MADF domains of *D. melanogaster Hmr* have a range of ionic properties different from the canonical MADF domain, and one MADF domain even has ionic properties inconsistent with DNA binding. This apparent functional plasticity is not restricted to *D. melanogaster* as *Hmr* orthologs from different species have evolved MADF domains with unique ionic properties (Table 2.1). Most striking is *D. mojavensis Hmr* in which three out of the four MADF domains have a net negative charge, making them more similar to the chromatin-binding SANT domain than to the canonical DNA-binding MADF domain.

Additionally, *Hmr* from *D. yakuba* and *D. santomea* encodes a microsatellite-like tandem repeat within its coding region that has undergone variable length expansion. Although many molecular evolutionary analyses by necessity ignore variable-length repeats because they introduce gaps into multi-alignments, they may be of functional importance. Studies have linked length variation of simple amino-acid repeats to evolutionary agility, meaning the capacity to generate a phenotypic range. One example found a correlation between morphological changes among dog breeds and repeat-length variation in two developmental regulatory genes, *Alx-4* and *Runx-2* [161]. Another example is the

well-studied clock gene *period* (*per*), which has a minisatellite-like coding repeat of alternating threonines (Thr) and glycines (Gly), the length of which shows a significant north-south cline across Europe that appears to have been maintained by natural selection [162, 163]. These examples justify further investigation of whether repeat-length variation in *Hmr* also contributes to its functional divergence.

The most striking observation from our analysis is that *Hmr* has undergone recurrent positive selection in multiple *Drosophila* lineages. We have also found that sequence evolution in *Hmr* has resulted in large variation in repeat-tract lengths among orthologs and has significantly altered the ionic properties of its MADF domains. Similar investigations of other HI genes will be necessary to address whether features such as recurrent selection and protein sequence plasticity are peculiar to *Hmr* or instead reflect general features of genes that cause reproductive isolation.

CHAPTER 3

ASYMMETRIC HYBRID LETHAL EFFECTS OF THE D-M GENE *LHR* ARE THE RESULT OF *CIS*-BY-*TRANS* REGULATORY DIVERGENCE OF A CONSERVED INTERACTION.

3.1 Introduction

New species are isolated from one another by a variety of reproductive barriers. One widely observed barrier is Hybrid Incompatibility (HI), the inviability and/or sterility of interspecies offspring. The key premise of the Dobzhansky-Muller (D-M) model explaining the evolution of HI is that genetic changes fixed in one population need not be compatible with changes fixed in a different population [73, 74]. In the simplest case, two independently evolving lineages diverge from the ancestral state *aabb*, and fix new alleles arriving at the states *AAbb* and *aaBB*, respectively. Hybridization between the two generates a genotype unscreened by natural selection, *AaBb*, and incompatible interaction between the derived alleles *A* and *B* may cause developmental breakdown. Questions fundamental to understanding speciation then are: What molecular divergence between the ancestral and derived alleles is causing HI? Is this divergence at the level of regulatory or structural changes? What are the evolutionary forces causing this divergence?

Hybrids are the sum of two independently evolving genomes and thus suffer from multiple suboptimal interactions. For example, species-specific divergence at *cis* and *trans*-regulatory elements is associated with widespread transcriptional dysregulation in hybrids [60, 62]. However, despite this background of complex changes, singe loci that have a major effect on the HI phenotype

have been identified and several permutations of the D-M model have been revealed. These range from bi-allelic single locus incompatibilities to complex multi-locus incompatibilities [76, 83, 84]. Furthermore D-M interactions are not limited to derived alleles, as ancestral-derived incompatibilities have also been identified [78]. D-M loci also belong to diverse functional classes, ranging from mitochondrial splicing factors in yeast, cell wall protease in rice to nucleopore subunits in Drosophila [78, 76, 88, 89]. However, one unifying emerging trend is that divergence at the HI locus is frequently driven by natural selection [164]. This is exciting, because if molecular divergence created by selection is causing HI, then the phenotypic target of selection is, at least in part, the evolutionary basis of speciation. A major goal then is to understand the role of selection in the evolution of incompatible divergence.

Locating the phenotypic target of selection is complicated because HI genes are frequently pleiotropic, such as the *Nups* [107], or novel genes with unknown functions, such as *Hmr* [7, 5]. Moreover, molecular signatures of selection do not necessarily imply advantageous ecological adaptations; they could instead be the legacy of genetic conflicts such as between host and pathogen or the invasion of selfish genetic elements (reviewed in [165]). Strikingly, mechanistic studies on several recently characterized HI genes implicate heterochromatin-associated functional divergence as the cause of incompatibility. *OdsH* is a fast-evolving homeodomain protein that mislocalizes to the heterochromatic Y-chromosome in hybrids [85]; *Zhr* is a species-specific repetitive sequence, and hybrid lethality is caused by improper segregation of this heterochromatic satellite [111]. Although, it is unclear why mislocalization of *OdsH* causes sterility and why the *Zhr* satellite is unstable, these observations have raised the possibility that their evolutionary origin lies in genetic conflict with selfish heterochro-

matic elements. The hybrid lethality gene *Lethal hybrid rescue* (*Lhr*) is another example of a rapidly diverging heterochromatin-associated HI gene [51]. It is an ideal candidate for testing the generality of the conflict-driven HI hypothesis.

Crosses between *D. melanogaster* females and sibling-species males result in lethal hybrid sons and sterile hybrid daughters. The incompatible D-M interaction in hybrid males can in part be explained by epistatic interaction between two genes, *Hybrid male rescue* (*Hmr*) on the *D. melanogaster* X-chromosome and *Lhr* on the *D. simulans* 2^{nd} chromosome [50, 51]. A loss of function mutation in either *Hmr* or *Lhr* alone is sufficient to suppress the lethality of hybrid sons [130, 166]. Thus it is the wild type activity of these genes that causes hybrid breakdown. Consistent with the expectation of functional divergence, the rescue of hybrid lethality via *Lhr* is asymmetric; removal of *simulans Lhr* rescues lethal hybrid sons but removal of *melanogaster Lhr* does not [51]. Moreover, the *Lhr* protein coding sequence (CDS) has diverged extensively under selection. Therefore it was surprising when a recent study comparing *Lhr* orthologs failed to detect any differences in their hybrid lethal activity or heterochromatin-association [86]. One explanation for these unexpected results is that an artificial over-expression system was used for genetic assays. Based on studies of *Hmr* [50] and also on results presented here, hybrid lethality appears to be highly sensitive to gene dosage effects, suggesting the possibility that overexpression might mask the evidence for functional divergence. Secondly, LHR localization was assayed using polytene chromosomes, which have relatively limited resolution for heterochromatin.

In order to uncover the functional divergence underlying asymmetric rescue

properties of *Lhr* orthologs, I used a native-promoter driven transgenic system that allows a sensitive comparison of the functions and localization properties of *D. simulans* and *D. melanogaster Lhr* orthologs. Here I assay and compare *Lhr* function in both pure species and hybrids using three sets of experiments: (1) genetic tests for hybrid lethal activity and interaction with its D-M partner, *Hmr*; (2) detailed cytological mapping of the heterochromatic localization of LHR and (3) expression analysis comparing transcriptional levels of the *Lhr* orthologs.

3.2 Materials and Methods

3.2.1 Drosophila stocks

All crosses were done at room temperature or at 18°C where explicitly stated. At least 2 replicates were done for each cross. Each interspecific cross was initiated with ≈15-20 1 day old *D. melanogaster* virgin females and ≈30-40 3-4 day old sibling species males. The nomenclature used for the transgenic lines and a complete description of the constructs used to generate them are included in Table 3.1.

3.2.2 DNA Constructs

To make a modified pCasper4 containing the attB site, I PCR amplified a 280bp fragment using the pTA plasmid (gift from Michele Calos) as the template. This PCR product, along with flanking *Sal*I sites was cloned into the compatible *Xho*I site of pCasper4 to create the plasmid pCasper4\attB. I chose to construct *Lhr*

Table 3.1: Transgene nomenclature

	Full name of construct	**Transgene short-hand**
1	$P\{w^{+mC}$ *Bap55*$^{t4.8}$ *Lhr*$^{t4.8}$ = *mel-Lhr*$\}$	$\Phi\{$*mel-Lhr*$\}$
2	$P\{w^{+mC}$ *Dsim\Bap55*$^{t4.8}$ *Dsim\Lhr*$^{t4.8}$ = *sim-Lhr*$\}$	$\Phi\{$*sim-Lhr*$\}$
3	$P\{w^{+mC}$ *Bap55*$^{t4.8}$ *Lhr::HA*$^{t4.8}$ = *mel-Lhr-HA*$\}$	$\Phi\{$*mel-Lhr-HA*$\}$
4	$P\{w^{+mC}$ *Bap55*$^{t4.8}$ *Lhr::YFP*$^{t4.8}$ = *mel-Lhr-YFP*$\}$	$\Phi\{$*mel-Lhr-YFP*$\}$
5	$P\{w^{+mC}$ *Dsim\Bap55*$^{t4.8}$ *Dsim\Lhr::HA*$^{t4.8}$ = *sim-Lhr-HA*$\}$	$\Phi\{$*sim-Lhr-HA*$\}$
6	$P\{w^{+mC}$ *Bap55*$^{-}$ *Lhr::HA*$^{t4.8}$ = Δ*Bap55 mel-Lhr-HA*$\}$	$\Phi\{\Delta$*Bap55 mel-Lhr-HA*$\}$

Table 3.2: Primers used in the materials and methods

No.	Sequence	Capitalized region
691	tactatAAGCTTtggttgttccacacgactttatcg	HindIII
664	tcgcatAAGCTTctggcaggtggtaaccgatacgg	HindIII
597	gggtttGCGGCCGCttccacacgactttatcgacagga	NotI
598	gggtttGGATCCcggctcctcaaacattcctttatg	BamHI
728	TGCATAGTCCGGGACGTCATAGGGATA GCCCGCATAGTCAGGAACATCGTATGG GTACATtgttctcagcgtaggccg	3xHA tag
729	CCCTATGACGTCCCGGACTATGCAGGA TCCTATCCATATGACGTTCCAGATTACG CTtgactttctttcgtataaaatgc	3xHA tag
730	tgttctcagcgtaggccgccttgagc	
731	TCAAGGCGGCCTACGCTGAGAACAaagg	YFP tag
732	atgtgcattttatacgaaagaaagTCACGTGGACCG GTGCTTGTACAGC	YFP tag
733	TACAAGCACCGGTCCACGTGActttctttcgta taaaatgcacataag	YFP tag
1171	gcaccatgTAATGACctatatggcggcgacgagat	Stop codons
1172	ccatatagGTCATTAcatggtgccgccactcat	Stop codons
1086	gtcgcccacatgacacaag	
1087	ctctttgcaaggcattacatctg	
949	gtcgacgatgtaggtcacggtc	
1177	gcagccgaacgaaattaaaa	
1147	gcatttagaagtggaggtcctcg	
1148	gctctttcacccgtatcgctttaag	

transgenes with *Lhr* under the control of its native regulatory sequences. I used a 4.8kb genomic fragment that spans 2.7kb upstream and 1kb downstream of the *Lhr* CDS. This fragment includes the complete CDS of the adjacent gene *Bap55* (Figure 3.1).

To generate the p{sim-Lhr} construct I amplified a 4.8kb fragment from *D. simulans* w^{501} genomic DNA, using primer pairs 691/ 664. This PCR product was gel purified and cloned into the pCR-BluntII TOPO vector (Invitrogen), according to manufacturer's directions. The insert was sequenced completely and subcloned into the MCS of pCasper4\attB using *Not*I and *Kpn*I restriction enzymes.

The p{mel-Lhr} construct was generated similarly, a 4.8kb fragment was PCR amplified from wild type *D. melanogaster* (strain Canton-S) genomic DNA using primer pairs 597/ 598, and TOPO cloned into pCR-BluntII vector. The forward primer contains a *Not*I site, allowing the insert to be released as a *Not*I fragment and cloned into the *Not*I site of pCasper4\attB. A clone was chosen with orientation identical to that in p{sim-Lhr}.

To construct p{sim-Lhr-HA} a triple-HA tag was added in-frame to the C-terminus of *Lhr* CDS using a two-piece fusion PCR strategy. The two overlapping PCR products were amplified using p{sim-Lhr} as the template, with primer pairs 691/728, and 729/664. These fragments were used as templates for the fusion PCR, and the gel-purified product was TOPO cloned into the pCR-BluntII vector and sequenced completely. The insert was then subcloned into pCasper4\attB exactly as in p{sim-Lhr}. The construction of p{mel-Lhr-HA} followed the same logic, using the primer pairs 597/728 and 729/598.

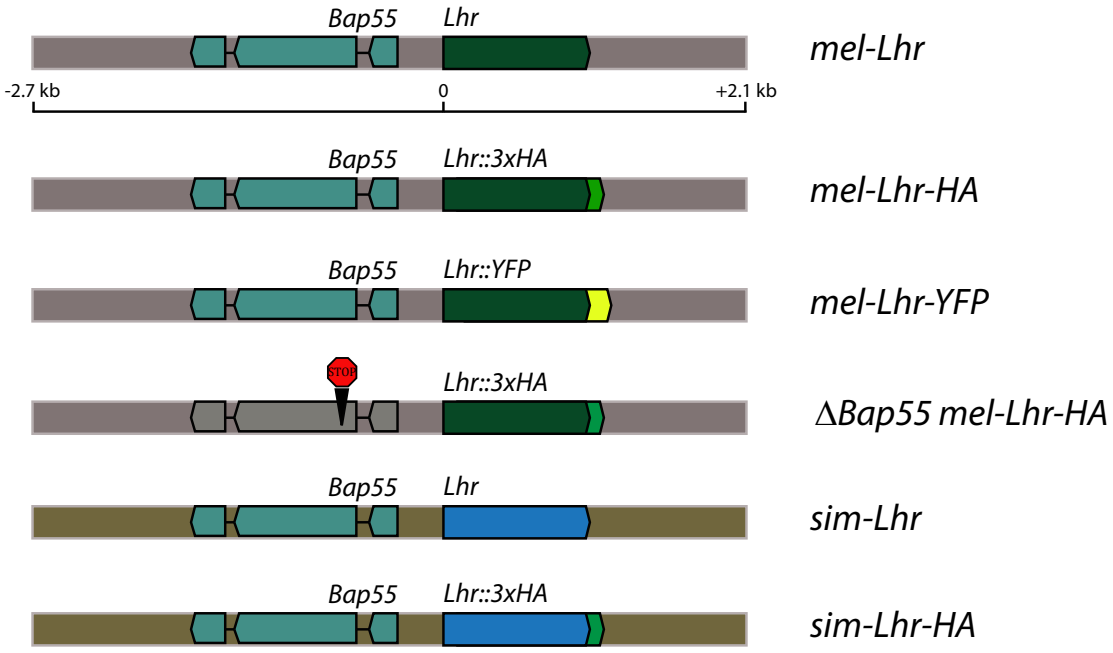

Figure 3.1: A schematic of the *Lhr* constructs. All constructs contain the full *Lhr* and *Bap55* coding sequences and UTRs. The "stop" in Δ*Bap55 mel-Lhr-HA* represents the insertion of "TAA TGA C", i.e. two stop codons and a frame shift mutation. The *mel-Lhr* construct is drawn to proportion, the HA and YFP epitope tags in other constructs are not drawn to proportion.

To synthesize the p{mel-Lhr-YFP} construct a three-piece fusion PCR strategy was used, the first and last PCR products, containing upstream and downstream genomic regions respectively, were amplified using p{mel-Lhr} as the template, with primer pairs 597/730 and 733/598. The central PCR product containing the YFP-tag was amplified from p{w^{+mC} UASLhr::Venus=UAS-Lhr::YFP} [51], with primer pair 731/732. The 3 overlapping PCR products were used as templates for the fusion PCR, and cloned into the pCR-BluntII vector and sequenced completely. The insert was subcloned into pCasper4\attB exactly as in p{mel-Lhr}.

The p{ΔBap55 mel-Lhr-HA} construct is identical to p{mel-Lhr-HA} except that the *Bap55* CDS is interrupted by the insertion of "TAA TGA C", i.e. two stop codons and a frame shift mutation after the second methionine at position 6. Two overlapping PCR products were amplified using p{mel-Lhr-HA} as template, with primer pairs 597/1171 and 1172/598. The products were stitched together using fusion PCR and cloned into pCasper4\attB exactly as done in p{mel-Lhr}.

3.2.3 Transgenic fly lines

ϕC31-mediated transformants of *D. melanogaster* were performed by Genetic Services. The integrations sites used were: i) P{CaryP}attP2 [167] and ii) M{3xP3-RFP.attP}ZH-86Fb [168] at cytological positions 68A4 and 86Fb, respectively. Site-specificity of intergration was tested using the PCR assays of Venken et. al. (2006) [169]. I also developed attP docking-site specific PCR assays, primer pairs 1086/1087 for attP2, and 949/ 1177 for ZH-86Fb. All *D. melanogaster* trans-

formants were crossed into the strain w^{1118}. Random P-element mediated integration was used to transform the *D. simulans* w^{501} strain with p{sim-Lhr-HA}.

3.2.4 Quantitative RT-PCR

Total RNA was isolated using the Trizol Reagent (Invitrogen), followed by DNaseI (Roche) treatment and purification using RNeasy columns (Qiagen). First strand cDNA was synthesized from 4μg of total RNA using the SuperScriptIII first-strand synthesis system (Invitrogen) with the oligo$(dT)_{20}$ primer in a 20μl reaction according to the manufacturer's instructions. Quantitative real time PCR (rt-qPCR) was performed on a Biorad MyiQ cycler with SYBR detection using the 2x supermix from Biorad. Relative concentrations of *Lhr* transcripts were calculated against *rpl32* as the reference gene with *rpl32* primers from Fiumera et. al. (2005) [170]. For *Lhr* a primer pair (1147/1148) was developed that recognized conserved sequences and that amplified both *D. melanogaster* and *D. simulans Lhr* with equal and high efficiency. For each sample real-time PCR on test and reference genes was done in technical triplicates, and the standard-curve method was used to estimate transcript abundance. For each genotype RNA was isolated from between 3 and 4 independent 6-10 hr-old embryo collections. For all genotypes except *D. simulans P{sim-Lhr-HA}* cDNA was synthesized twice from each RNA isolate.

3.2.5 Pyrosequencing

RNA was extracted from 3-5 day-old larvae collected from non-crowded vials. In hybrid crosses the *D. melanogaster* mothers carried the X-linked mutation y^-,allowing one to determine the sex of larvae using mouth hook coloration (daughters are y^+ and sons y^-). Total RNA and genomic DNA were simultaneously extracted from the same biological samples using the SV RNA system (Promega). For the pure species control, RNA and genomic DNA were extracted once from a single biological collection, followed by a single round of cDNA synthesis. For the hybrid samples, RNA and genomic DNA were extracted from four independent biological samples. cDNA was synthesized twice from each independent RNA isolate. Pyrosequencing measurements were performed in triplicate on each cDNA and in duplicate on each genomic DNA.

3.2.6 Western Blotting

Whole cell extracts were obtained by grinding samples in approximately three times the volume of lysis buffer (50 mM Tris-HCl pH 7.5, 10 mM EDTA, 1.25% TritonX-100, 1X Roche protease inhibitor tablet). Extracts were cleared by centrifugation at 14,000 rpm for 10 min at 4°C. Total protein concentration of the cleared extracts was measured using Bradford assay (Biorad) and the samples were boiled in 0.5X volume of 4X SDS-Sample buffer. For most westerns 40μg of total protein was loaded in each lane. Primary antibodies used were: rat anti-HA 3F10 (Roche, 1:1000) and mouse anti-tubulin T5168 (Sigma, 1:10,000). HRP conjugated goat anti-rat and goat anti-mouse secondary antibodies (Jackson, 1:5,000) were used and detected with ECL Western blotting substrate (Pierce).

3.2.7 FISH and Immuno-staining

Embryo FISH and immuno-FISH was performed as in Ferree et. al. (2009) [111] and immunostaining of ovarioles was performed as in Aruna et. al. (2009) [5] with the following antibodies: Rat anti-HA 3F10 (Roche; 1:100), mouse anti-HP1 C1A9 (DSHB; 1:100), rabbit anti-histone H3 lysine 9 dimethylation (Upstate 07-441; 1:100), rabbit anti-Cid (a gift from S. Henikoff; 1:1000), rabbit anti-GFP (Abcam ab6556; 1:300), mouse anti-Fibrillarin (Cytoskeleton Inc. AFb01; 1:400) and mouse anti-Hts 1B1 (DSHB; 1:4). DNA was stained using TOPRO-3 iodide (Molecular Probes) or Vectashield containing DAPI (Vector Laboratories). All imaging was conducted at the Cornell University Core Life Sciences Microscopy and Imaging Facility, using either a Leica DM IRB confocal microscope or an Olympus BX50 epifluorescent microscope, except for embryo images with a DAPI channel which were taken in the Plant Cell Imaging Center at the Boyce Thompson Institute, with a Leica TCS SP5 confocal microscope. Images were processed using Photoshop (Adobe, version 7.0). Contrast and brightness changes, when used, were applied globally across the image.

Quantification of dodeca signal in interphase larval brain tissue was done using ImageJ [171]. Watershed segmentation was applied on the DAPI-channel to generate a mask of nuclear territories. The Analyze Particle function was then used to identify individual nuclei as ROIs (regions of interest) and screened to exclude aberrant nuclear segmentations and non-nuclear entities. Each ROI was individually selected on the dodeca FISH channel of the same image and the FociPicker3D plug-in was used to identify regions of local maxima. I then calculated two measures to estimate the nuclear dispersion of dodeca satellite: (1) the total number of foci per nucleus and (2) the fraction of total nuclear area

occupied by the dodeca signal.

3.3 Results

3.3.1 Both *D. simulans* and *D. melanogaster Lhr* orthologs have hybrid lethal activity.

I used the ϕC31 site-specific integration system to generate parallel strains of *D. melanogaster* containing either *D. simulans* or *D. melanogaster Lhr* transgenes. Each *Lhr* ortholog was C-terminally tagged with HA and was expressed under the control of its native regulatory sequences (Figure 3.1). The transgenic constructs contained the eye-color marker *white*$^{+}$ and were each integrated into the attP2 site on the third chromosome. I tested the transgenes for wild type activity by assaying for complementation of the *D. simulans Lhr*^{1} hybrid rescue mutation. *D. simulans Lhr*^{1} is a loss-of-function mutation that acts as a dominant suppressor of hybrid lethality [166]. Complementation here means that the transgene provides sufficient wild type *Lhr* activity to suppress rescue by the *Lhr*^{1} mutation.

Complementation tests were performed by crossing *D. melanogaster* mothers heterozygous for the *Lhr-HA* transgene to *D.simulans Lhr*^{1} fathers. This cross generates two classes of hybrid sons: the control class that lacks the transgene and has white eyes, and the experimental class that inherits the transgene and has orange eyes. Complementation is detected as the lethality of orange-eyed sons. If hybrid lethal activity partitions discretely between *Lhr* orthologs, as

expected from the functional divergence interpretation of genetic asymmetry, sons inheriting the ϕ\{Dsim\Lhr-HA\} transgene should be lethal, while those inheriting ϕ\{Dmel\Lhr-HA\} should be viable.

Surprisingly both transgenes fully complemented the *D. simulans Lhr*^{1} mutation. I recovered no orange-eyed sons in either cross, which argues that both *D. simulans* and *D. melanogaster Lhr* orthologs have hybrid lethal activity (Table 3.3, crosses 1 thru 4). As this result was contrary to expectation I tested several possible causes of artifacts in the transgenic constructs. First, the C-terminal HA-tag does not affect *Lhr* function because untagged versions of both *mel-Lhr* and *sim-Lhr* also complement *Lhr*^{1} (Table 3.3, crosses 5 and 6). Second, the adjacent gene *Bap55* present in these constructs is not responsible for complementation because a modified *mel-Lhr-HA* transgene, ϕ\{ΔBap55 Dmel\mel-Lhr-HA\}, in which the *Bap55* CDS is interrupted by two stop codons and a frameshift mutation, also complements *Lhr*^{1} (Table 3.3 cross 7). Third, the results are not caused by other unknown aspects of the strain background or the attP2 site because the attP2 site itself without an integrated transgene does not complement *Lhr*^{1} (Table 3.3 cross 8). Furthermore *mel-Lhr-HA* integrated into a different site (attP86Fb) also complements *Lhr*^{1} (Table 3.3, cross 4). Fourth, these results are not due to an over-expression artifact because data presented below demonstrate that the *mel-Lhr-HA* transgene expresses *Lhr* at a level similar to the endogenous wild type locus (Figure 3.10B). These results clearly show that *D. melanogaster Lhr* has hybrid lethal activity even when expressed at its wild type level.

How does one reconcile these results with the original observation of a discrete functional divergence between *D. melanogaster* and *D. simulans Lhr*? Those

Table 3.3: *D. melanogaster* and *D. simulans Lhr* orthologs suppress hybrid rescue by *D. simulans Lhr*[1].

Cross	Transgenic contruct	*attP* site	Temp.	No. of hybrid females	No. of hybrid males	
					Genotype 1 +/+	Genotype 2 $\phi\{\}$/+
1	$\phi\{$*sim-Lhr-HA*$\}$	68A4	RT	232	92	0
			18° C	214	110	0
2	$\phi\{$*sim-Lhr-HA*$\}$	86Fb	RT	135	74	0
			18° C	100	57	0
3	$\phi\{$*mel-Lhr-HA*$\}$	68A4	RT	177	91	0
			18° C	240	122	0
4	$\phi\{$*mel-Lhr-HA*$\}$	86Fb	RT	263	121	0
			18° C	246	109	0
5	$\phi\{$*sim-Lhr*$\}$	68A4	RT	184	61	0
6	$\phi\{$*mel-Lhr*$\}$	68A4	RT	302	150	0
			18° C	217	84	0
7	$\phi\{\Delta$*Bap55, mel-Lhr-HA*$\}$	68A4	RT	324	156	0
			18° C	322	188	0
8	*P*$\{$*Casper4.attP,* $y^+\}$	68A4	RT	280	NA	160

Crosses were between *D. melanogaster* females *w;* $\phi\{\}$/+ heterozygous for the different transgenes tested in *D. simulans Lhr*[1] males. The transgenes carried a copy of the w^+ gene so the hybrid sons inheriting the transgene $\phi\{\}$/+ (genotype 2) were distinguished from their +/+ siblings (genotype 1) by their eye-color. The only exception was cross number 8 where *D. melanogaster* females homozygous for the integration-site without an inserted transgene were mated to *D. simulans Lhr*[1] males.

Table 3.4: A single dose of transgenic *mel-Lhr* is sufficient to suppress hybrid rescue by *D. simulans Lhr*[1]

	Hybrid females		Hybrid males			
	Df(Lhr)/+	*Bal/+*	*Df(Lhr)/+*		*Bal/+*	
Temp.			ϕ{*Lhr*}/+	+/+	ϕ{*Lhr*}/+	+/+
18 °C	124	173	0	67	0	62
RT	198	149	0	68	0	52

D. melanogaster females of the genotype *y,w; Df(2R)Lhr/CyO;* ϕ{*mel-Lhr-HA*}/+ were mated to *D. simulans Lhr*[1] males. Hybrids were scored as follows: *Df(2R)Lhr/+* progeny are CyO^+ and have straight wings; *Bal/+* progeny carry the *CyO* balancer chromosome and have curly wings. Hybrid male progeny that inherit the transgene are orange eyed, while the sibling brothers are white eyed. Hybrid female progeny with and without the transgene cannot be distinguished.

experiments were done in hybrid genotypes that had only a single dose of either *mel-Lhr* or *sim-Lhr* [51]. In contrast, the experiments here were performed by adding a transgenic copy of either *mel-Lhr* or *sim-Lhr* to hybrids that also carried the endogenous chromosomal copy of *mel-Lhr*. Since the results presented here suggest that *mel-Lhr* has hybrid lethal activity, increased dosage of *mel-Lhr* in the current experiments may explain why I have not observed a difference between the *mel-Lhr* and *sim-Lhr* transgenes. This hypothesis raises the question of whether the hybrid lethal activity of the ϕ\{Dmel\Lhr-HA\} transgene would be eliminated in a background lacking the chromosomal copy of *mel-Lhr*. To test this I crossed *D. melanogaster* mothers that were doubly heterozygous for ϕ\{Dmel\Lhr-HA\} and an Lhr^{-} deficiency to *D. simulans* Lhr^{1} fathers. If transgenic *mel-Lhr* behaves identically to the endogenous locus, then hybrid sons inheriting the Lhr^{-} deficiency along with the *mel-Lhr* transgene should be equivalent in *Lhr* dosage to $+/Lhr^{1}$ hybrid males and thus be viable. However, this was not the case as these hybrids were completely lethal (Table 3.4). This result indicates either that the ϕ\{Dmel\Lhr-HA\} transgene does not precisely phenocopy the native chromosomal *mel-Lhr* locus, or that the multi-locus Lhr^{-} deficiency has negative effects in hybrids caused by deleting flanking genes.

3.3.2 Interactions with *Hmr* reveal a difference in lethal activity between *sim-Lhr* and *mel-Lhr*.

Because of these unexpected results I used a more sensitive genetic assay to test for functional divergence between *mel-Lhr* and *sim-Lhr*. It had been previously demonstrated that *Lhr*-dependent hybrid lethality requires the presence of its

D-M partner, the *D. melanogaster* gene *Hmr* [51]. Although, hybrid sons rescued by the hypomorphic allele Hmr^{1} were made inviable by the GAL4-induced expression of *UAS-sim-Lhr*, the same *sim-Lhr* over-expression failed to complement rescue by the null allele $Df(1)Hmr^{-}$. These experiments were interpreted as demonstrating that *sim-Lhr* requires *mel-Hmr* to cause hybrid lethality.

I reasoned that the hypomorphic *Hmr* allele might exhibit sensitivities to the HI effect of the different *Lhr* alleles. I therefore introduced each of the *Lhr* transgenes into Hmr^{1} and $Df(1)Hmr^{-}$ mutant backgrounds and tested the effect of the transgenes on hybrid male viability in crosses to *D. mauritiana* and *D. simulans*. Crosses with the *sim-Lhr-HA* transgene recapitulated previous experiments: Hmr^{1} hybrid males carrying *sim-Lhr-HA* were essentially lethal at room temperature and strongly reduced in viability at 18°C, while $Df(1)Hmr^{-}$ hybrid males were equally viable with and without the transgene (Table 3.5). I then performed similar crosses with *mel-Lhr-HA*. This transgene had little effect on viability of $Df(1)Hmr^{-}$ males and the results were not significantly different compared to the crosses with *sim-Lhr-HA* (Table 3.5, 1 & 2). In crosses with the hypomorphic mutation Hmr^{1}, hybrids carrying *mel-Lhr-HA* had reduced viability compared to their non-transgene carrying siblings, particularly at room temperature. Strikingly, however, I found that in all four cross conditions the magnitude of the viability reduction was significantly less for *mel-Lhr-HA* compared to *sim-Lhr-HA* (Table 3.5, 3 & 4). These data demonstrate that *sim-Lhr* is more potent than *mel-Lhr* in creating the hybrid lethal interaction with *Hmr*, and that these *Lhr* transgenes thus do in fact demonstrate the expectation of functional divergence.

Table 3.5: *D. simulans Lhr* interacts more strongly with *Hmr* than *D. melanogaster Lhr*.

				Hybrid male genotype				
				D. mel Lhr		*D. sim Lhr*		
	Hmr alelle tested	Male parent	Temp.	+/+	$\phi\{\}$/+	+/+	$\phi\{\}$/+	Fisher's exact test P
1	*Df(1)Hmr*	*D. sim*	RT	128	160	278	288	0.2178
	Df(1)Hmr	*D. sim*	18°C	n.d.	n.d.	29	18	n.d.
2	*Df(1)Hmr*	*D. mau*	RT	124	195	119	124	0.02024*
	Df(1)Hmr	*D. mau*	18°C	140	120	50	45	0.904438
3	Hmr^1	*D. sim*	RT	181	33	35	0	0.00654**
	Hmr^1	*D. sim*	18°C	349	258	502	82	0.00000***
4	Hmr^1	*D. mau*	RT	351	117	159	2	0.00000***
	Hmr^1	*D. mau*	18°C	497	388	476	256	0.00029***

Two *Hmr* alleles were tested for interaction with transgenic *Lhr* orthologs: a null and a hypomorph, specified in the table as *Df(1)Hmr* and Hmr^1 respectively. Crosses were set up with *D. melanogaster* females that carried an *Hmr* mutant allele and were heterozygous for *Lhr* transgene, denoted as $\phi\{\}$ in the table. Complete genotype of the females is as follows: null mutation, *y w Df(1)*Hmr^- *v/ FM6 (B, dm, sc, y, w);* ϕ*{transgene,* w^+*}/+* and hypomorph, *w* Hmr^1 *v;* ϕ*{transgene,* w^+*}/+*. Each *D. melanogaster* female genotype was mated to males from two sibling species, *D. simulans 9-29v* and *D. mauritiana Iso105*. Hybrid male progeny that inherit the transgene were identified as orange eyed, while the sibling brothers were white eyed. The FET is comparing the relative viability of hybrid sons inheriting the *D. melanogaster Lhr* transgene vs the relative viability of sons inheriting the *D.simulans Lhr* transgene in parallel crosses.

3.3.3 LHR partially localizes to the dodeca satellite within heterochromatin.

Coding sequence evolution leading to different localization patterns is one possible cause of *Lhr* functional divergence. In order to test this hypothesis I examined the cellular localization of LHR orthologs in their wild type backgrounds using the *Lhr* transgenes. In *D. melanogaster* LHR protein is most abundant during embryogenesis (Figure 3.2A). I next analyzed the distribution of LHR through the cell cycle and found a cyclical on-off pattern, with localization to chromatin mainly during interphase (Figure 3.3). This pattern is identical to its interaction partner, Heterochromatin Protein 1 (HP1) [172].Thus, I focused on interphase nuclei, and unless otherwise specified all images were taken at embryonic nuclear cycles 12-14, when heterochromatin is first observed. Consistent with previous results, LHR-HA co-localized with HP1 at DAPI-rich heterochromatic foci on the apical surface of the nuclei (Figure 3.4A). Unlike HP1, however, which is found throughout the nuclear compartment including euchromatin, LHR is restricted to heterochromatin. Consistent with being localized to a sub-domain of HP1, LHR localization strongly overlapped with Histone-3 lysine 9 dimethylation (H3K9me2), a histone modification specific to pericentric heterochromatin [173].

LHR staining was also observed in the embryonic germline precursors, the pole cells, and in the somatic and germline cells of the ovary (Figure 3.4B), where it again co-localized with H3K9me2 (Figure 3.4C). However, LHR was clearly excluded from the nucleolus, a sub-compartment within heterochromatin consisting of rDNA repeats (Figure 3.4C). This suggested that LHR has a specific distribution within heterochromatin. I therefore used immuno-FISH to

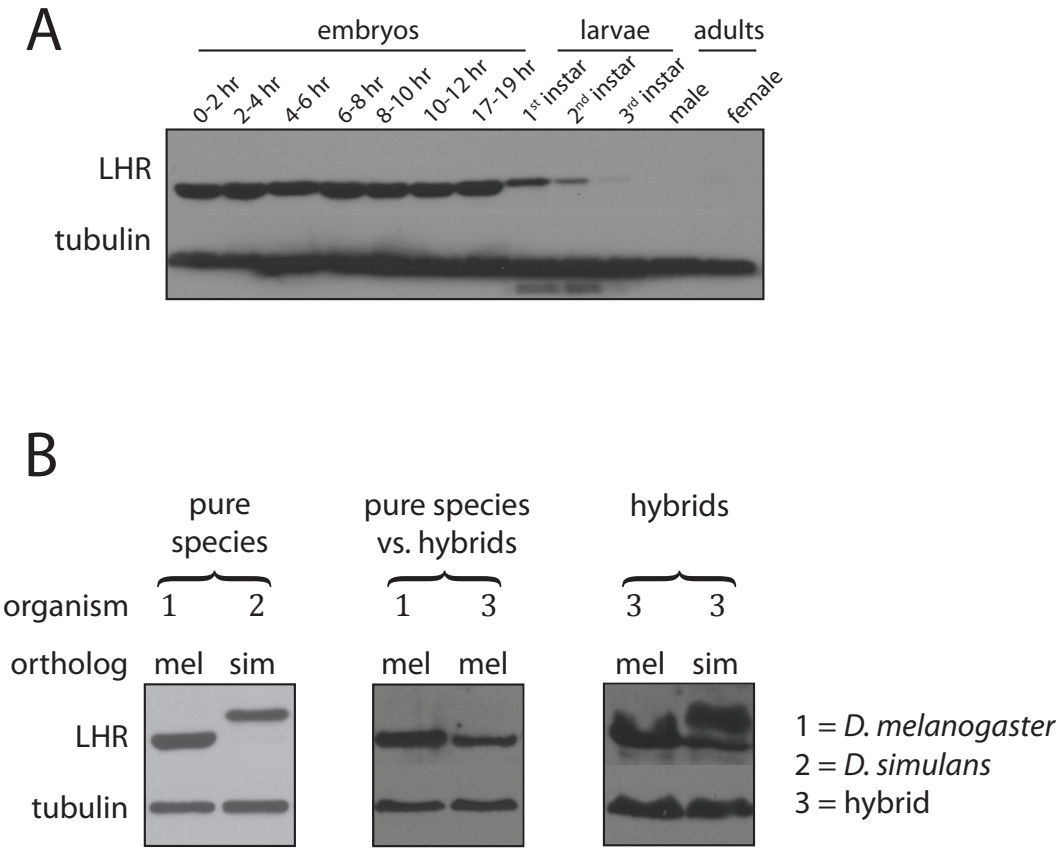

Figure 3.2: Comparison of LHR protein across developmental stages and genetic backgrounds using Western blots.(A) A developmental time course of mel-LHR-HA protein from a transgene homozygous in *D. melanogaster*. (B) LHR protein in 0-16 hr embryos from HA-tagged transgenes in different genetic backgrounds. (left panel) Two copies of *Lhr-HA* transgenes in their own species; (middle) One copy of the *mel-Lhr-HA* transgene in *D. melanogaster* and in hybrids; (right) One copy of the *D. melanogaster* and *D. simulans Lhr-HA* transgenes in hybrids. Immunoblots were hybridized with anti-HA antibodies, and anti-tubulin was used as a loading control. The molecular weight of *D. simulans* LHR is predicted to be 2 kD greater than *melanogaster* LHR, thus the size shift between the two orthologs.

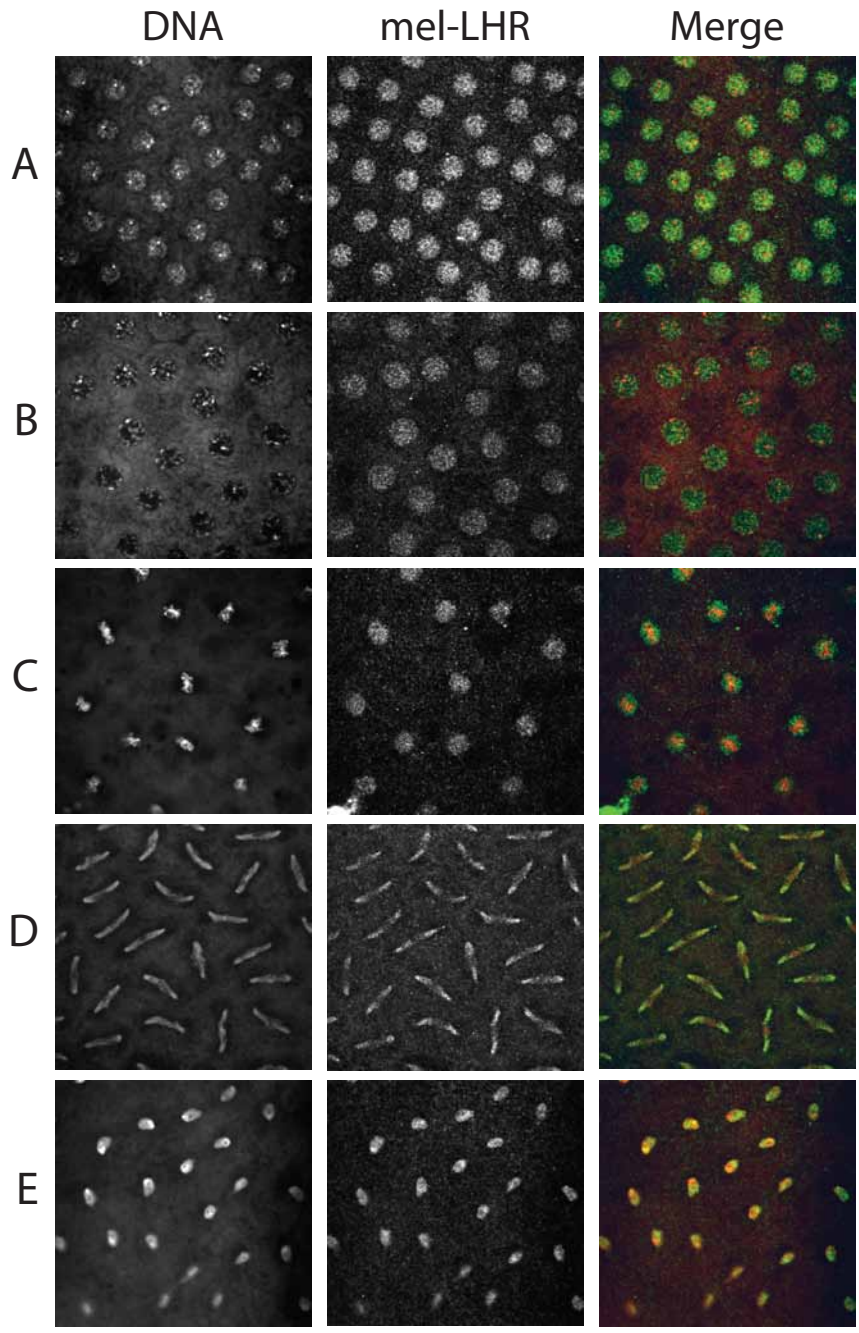

Figure 3.3: LHR distribution through the cell cycle. mel-LHR-HA detected with anti-HA (green) and DNA stained with TOPRO-3 (red) in *D. melanogaster* nuclear cycle 10 embryos during: (A) interphase; (B) prophase; (C) metaphase; (D) anaphase; (E) telophase.

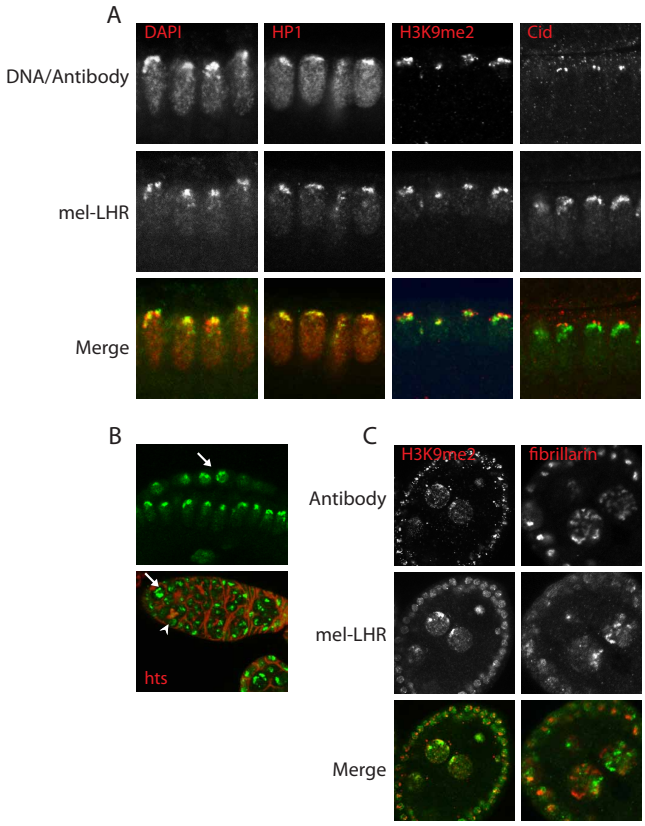

Figure 3.4: *D. melanogaster* LHR localizes to pericentric heterochromatin.(A) *D. melanogaster* cycle 14 embryos co-stained for mel-LHR-HA (green) and different heterochromatic markers (red). mel-LHR-HA localizes as distinct foci within heterochromatin marked by DAPI and anti-HP1, and shows colocalization with pericentric heterochromatin marked by anti-H3K9me2. mel-LHR-HA does not colocalize with centromeres as marked by anti-Cid . (B) Top, the posterior region of a cycle 12-14 embryo showing mel-LHR-HA (green) in the pole cells (germline precursor cells; arrow). Bottom, a germarium from a 2-3 day old female ovary with the anterior end to the left. mel-LHR-HA is found in all germline cells including the stem cells (arrow) and the developing cysts (arrowhead). The fusome and follicle cell membranes are marked using anti-hts (red). A portion of an egg chamber is visible in the bottom right. (C) Stage 4-6 egg chambers from 2-3 day-old *D. melanogaster* ovaries, stained for mel-LHR-HA (green) and heterochromatic markers (red). mel-LHR-HA is found in both the polyploid germline nurse cells (large cells in the centre) and in somatic follicle cells (small surrounding cells), and colocalizes to a sub-compartment of heterochromatin marked by anti-H3K9me2 but not to the nucleolus, stained with anti-fibrillarin.

investigate the localization pattern of LHR relative to various pericentric satellites in *D. melanogaster*. I observed no overlap between LHR and the 359 bp satellite, a 4-5 Mb block on the X-chromosome [174, 175], nor between LHR and the highly abundant AATAT satellite, which is distributed across multiple chromosomes [176] (Figure 3.5A & B). In contrast, LHR consistently overlapped with dodeca, a G.C-rich pericentric satellite on the third chromosome [177], although a substantial amount of LHR is also found in other heterochromatic regions that I have not mapped. During metaphase, however, the majority of LHR signal localized to four discrete foci along the metaphase plate. Strikingly, each LHR focus corresponded to the pericentric region of the third chromosome, as identified by overlapping dodeca signal (Figure 3.5B).

3.3.4 LHR heterochromatic localization between *D. melanogaster* and *D. simulans* is conserved

I next wanted to test whether LHR localization is conserved in *D. simulans*. I constructed transgenic lines of *D. simulans* using the same epitope tagged sim-LHR-HA construct that I described in the above genetic assays. Like mel-LHR, sim-LHR in *D. simulans* also localized to apical heterochromatic foci, as marked by DAPI (Figure 3.7A). Dodeca, which is present only on the pericentric region of the third chromosome in *D. melanogaster*, is present on the pericentric heterochromatin of both the second and the third chromosomes in *D.simulans* [178]. I confirmed this difference and mapped the dominant dodeca signal to the *D. simulans* second chromosome in mitotic brain squashes (Figure 3.6A). I also noted significant differences in the interphase organization of dodeca between

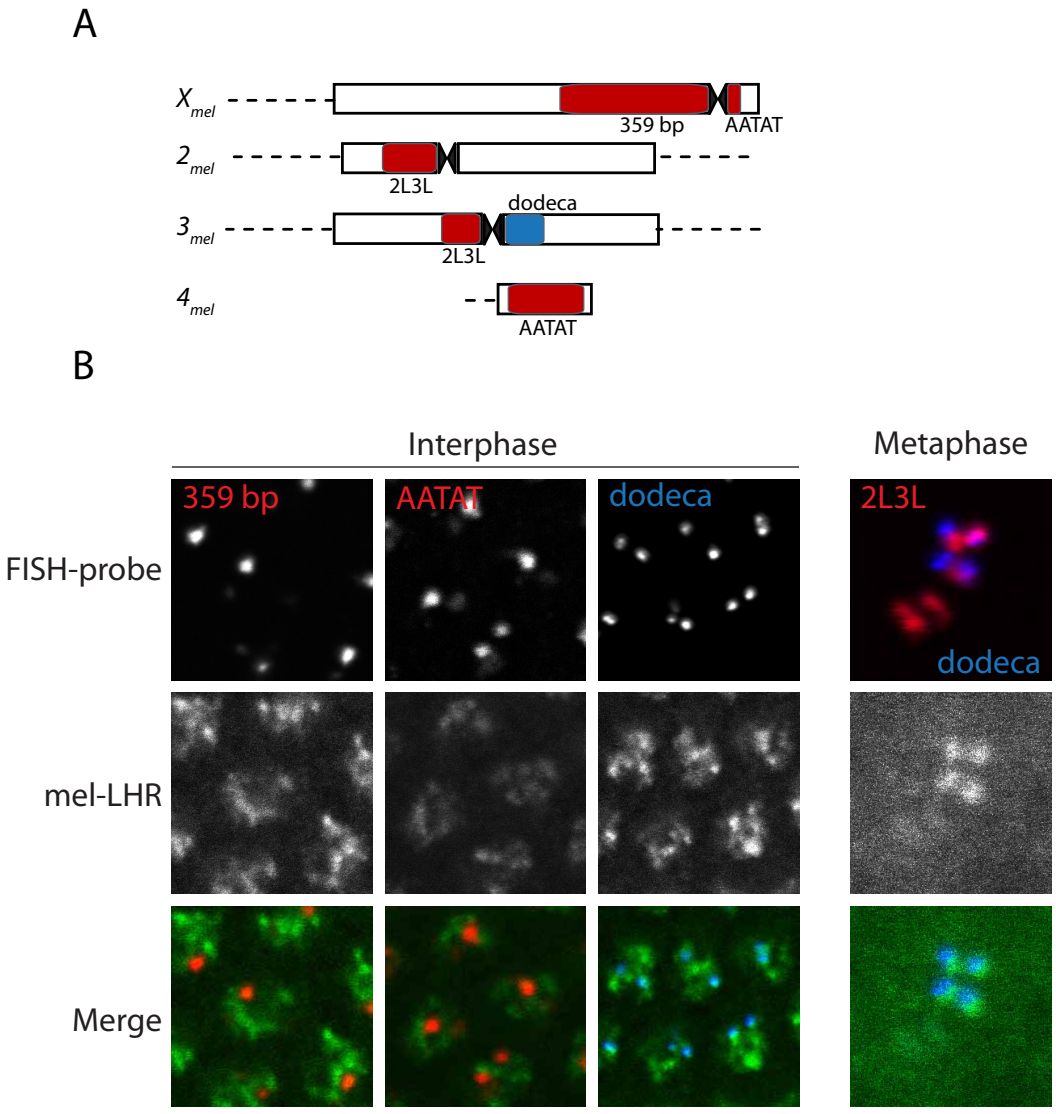

Figure 3.5: *D. melanogaster* LHR partially localizes with the dodeca satellite (A) Schematic of satellites used as targets of FISH. (B) immuno-FISH experiments in *D. melanogaster* embryos with anti-HA (green) to detect mel-LHR-HA and various FISH probes (red or blue). In interphase nuclei LHR shows no overlap with the 359 bp and AATAT satellites but partially co-localizes with the dodeca satellite. Right panel shows a mitotic nucleus with the pericentric regions of chromosomes 2 and 3 (marked by the 2L3L satellite) aligned at the metaphase plate. In the merge LHR signal is clearly present only at the 3^{rd} chromosome, marked by the dodeca satellite.

species. I quantified the number of dodeca foci per nucleus and the fraction of nuclear space occupied in interphase nuclei from wild type brains. The dodeca signal in *D. simulans* was broken up into more foci and occupied a greater nuclear volume, indicating that dodeca-containing heterochromatin had evolved species-specific nuclear organization properties (Figure 3.6B).

Remarkably, despite this divergence in both chromosomal location and structure of dodeca, immuno-FISH mapping in *D. simulans* showed that sim-LHR partially co-localized with dodeca in interphase nuclei (Figure 3.7A).It is highly unlikely that this conserved pattern is because LHR orthologs share a DNA-binding activity specific to the dodeca sequence for three reasons: (1) LHR CDS encodes no recognizable DNA-binding domain, (2) LHR localization to heterochromatin is dependent on HP1 binding [114, 86] and (3) LHR signal is neither restricted to dodeca nor perfectly overlapping with it (Figure 3.5B and 3.7). Thus, this shared localization to dodeca could also be interpreted as divergence of localization properties driven by co-evolution with species-specific heterochromatic sequences.

To distinguish between the two I examined the localization of LHR in a foreign species background. I did immuno-FISH experiments on the ϕ\{Dsim\Lhr-HA\} transgenic line, which expresses sim-LHR in *D. melanogaster*. I found that sim-LHR localized to the H3K9me2-enriched heterochromatic regions, and co-localized with the dodeca satellite in a pattern identical to that seen for mel-LHR (Figure 3.7B). In order to directly compare the localization of LHR orthologs within the same nucleus, I generated a recombinant transgenic line that expressed both YFP-tagged mel-LHR and HA-tagged sim-LHR. The two LHR orthologs showed complete overlap, demonstrating that the heterochromatic

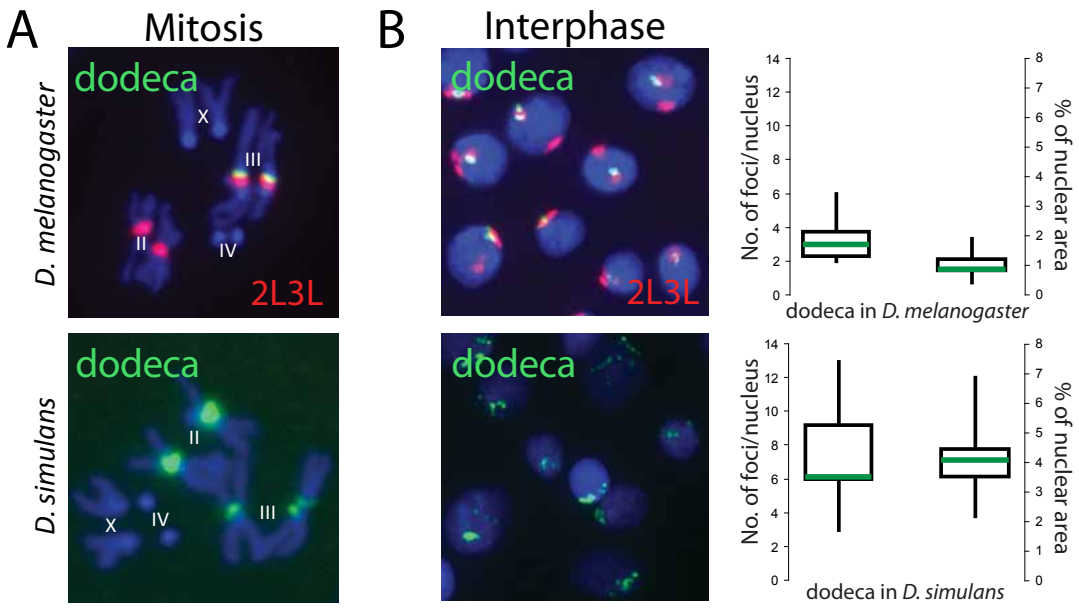

Figure 3.6: Divergent localization and interphase organization of the dodeca satellite between *D. melanogaster* and *D. simulans*. FISH to mitotic (A) and interphase (B) nuclei from 3rd instar larval brain cells with probes to dodeca (green) and 2L3L (red).Right panel shows quantification of the nuclear distribution of the interphase dodeca FISH signal. The mean value is indicated by the green line (n=10).Boxes span the interquartile range and whiskers extend to the maximum and minimum values.

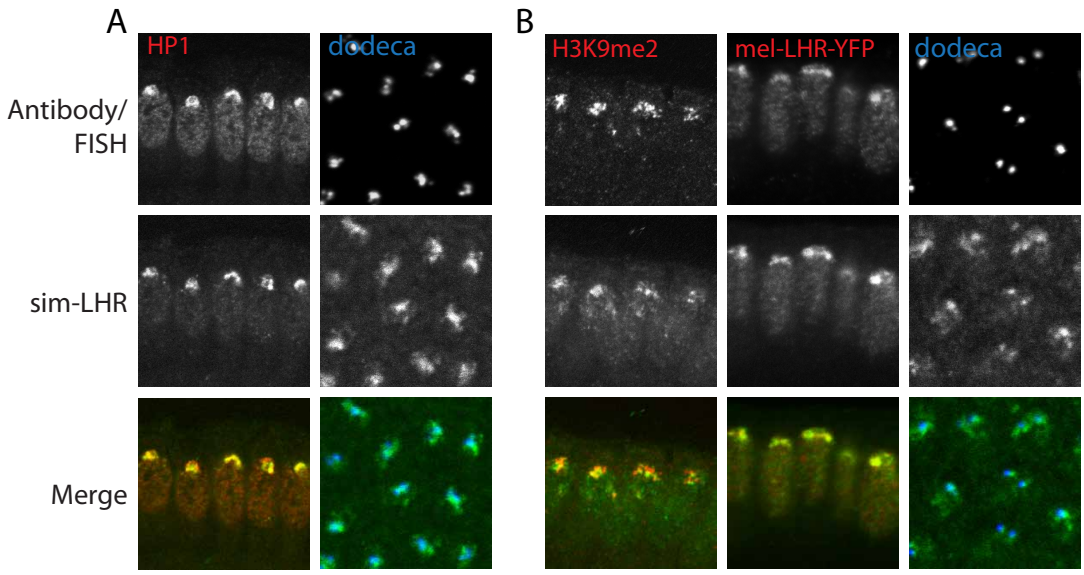

Figure 3.7: LHR orthologs have conserved heterochromatic localization properties.(A) *sim-Lhr-HA* transgene in *D. simulans* embryos.Left panel, Anti-HA (green) detects sim-LHR-HA co-localizing with HP1 (red) in the apical heterochromatin. Right panel, sim-LHR-HA partially co-localizes with dodeca satellite FISH (blue). (B) *sim-Lhr-HA* transgene in *D. melanogaster* embryos. sim-LHR-HA was detected with anti-HA and co-localizes with H3K9me2 and mel-LHR-YFP, detected with anti-GFP (red). sim-LHR-HA also partially overlaps with the *D. melanogaster* dodeca satellite.

localization properties of LHR orthologs are indeed conserved (Figure 3.7B).

3.3.5 Wild type heterochromatin and LHR localization in incompatible hybrids.

To determine whether general heterochromatin states are perturbed in hybrids I looked at HP1 and H3K9me2 localization. Although hybrid embryos were not sexed in this experiment, the staining appeared uniformly wild type. In order to specifically compare LHR and/or dodeca localization in hybrid males versus females, I developed a FISH probe that hybridized to the simulans Y-chromosome (Figure 3.8). I found that mel-LHR staining was enriched within the apical heterochromatin compartment in both sexes, and that it overlapped with the dodeca signal (Figure 3.9A). Importantly, I detected no difference in dodeca organization and LHR localization between lethal hybrid males and viable hybrid females (Figure 3.9B).

I then looked at heterochromatin states of hybrid larvae since developmental defects could become more apparent later in development. I examined several satellite classes in larval neuroblasts, because perturbations to the general heterochromatin structure are often reported as disruptions in the organization and clustering of satellite DNA repeats [180]. Consistent with the embryo staining, I saw no defects in the organization of dodeca and several other satellite classes in either lethal male or viable female larvae (Figure 3.9D). Moreover, despite obvious differences in the pericentric heterochromatic sequences between homologous chromosomes, somatic pairing was entirely unaffected in hybrid nuclei.

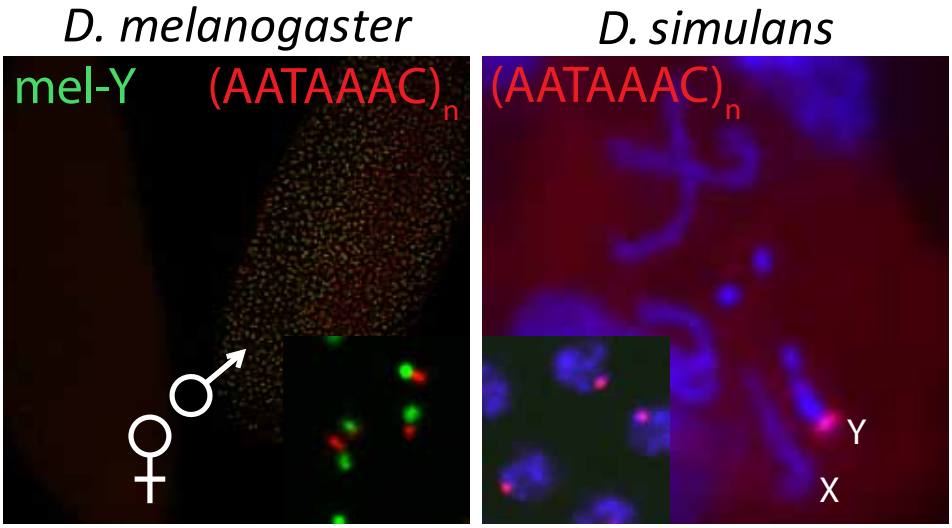

Figure 3.8: FISH mapping demonstrating that (AATAAAC)n is a *D. simulans* Y-specific satellite. The (AATAAAC)n was identified as a satellite sequence on *D. melanogaster* Y by Bonaccorsi & Lohe (1991) [179]. The left panel contains a *D. melanogaster* female and male embryo. Both the *D. melanogaster* Y-specific probe (AATAC)n and the *D. simulans* Y probe (AATAAAC)n hybridize specifically to the male embryo. The inset is a higher magnification of nuclei from the male embryo. The right panel is a mitotic spread from wild type *D. simulans* male 3rd instar larval brain cells. (AATAAAC)n hybridizes to a single region on the Y-chromosome. The inset is nuclei from a *D. simulans* male embryo, (AATAAAC)n signal is seen as a single dot in each nucleus.

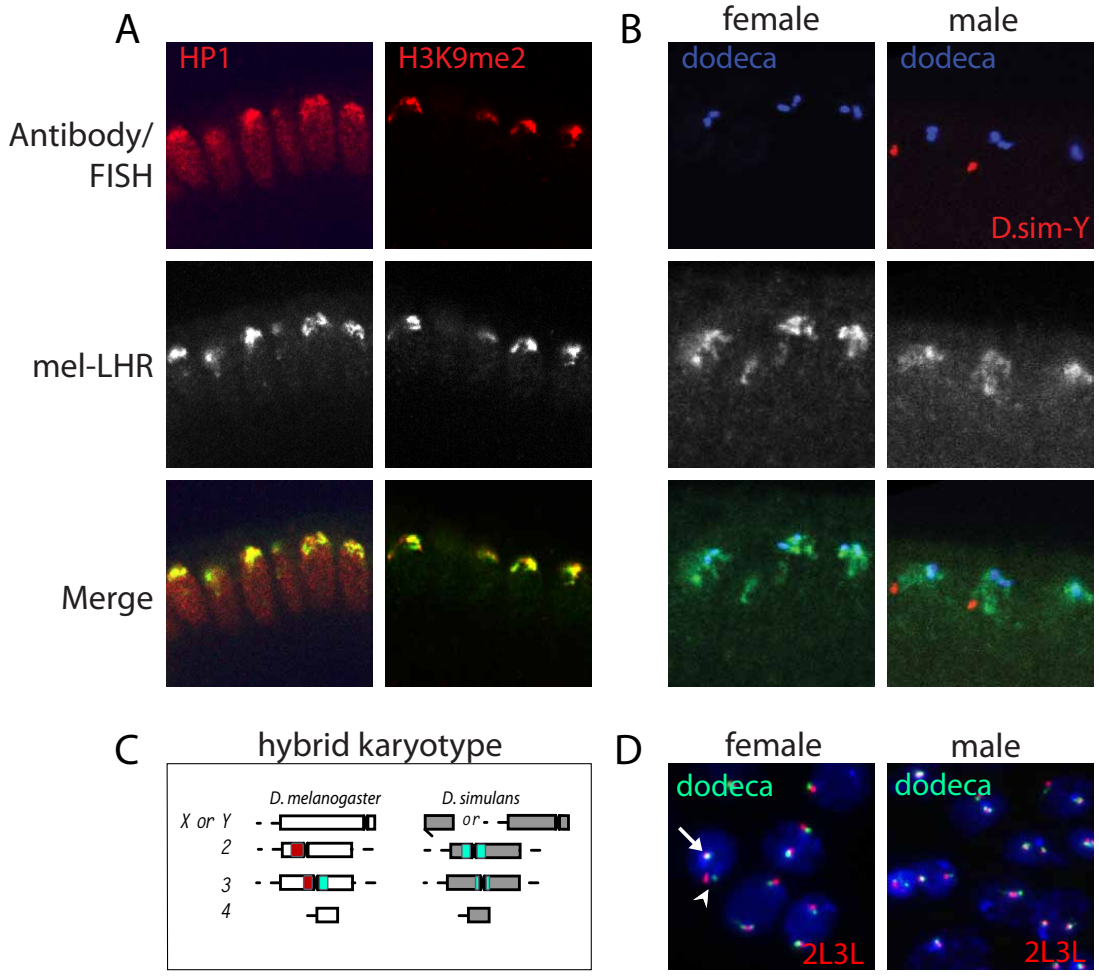

Figure 3.9: Normal LHR localization and organization of heterochromatin in hybrids. (A) mel-LHR-HA (green) colocalizes with HP1 and H3K9me2 (each red), similarly to wild type (see Figure 3.4). (B) mel-LHR-HA partially colocalizes to dodeca satellite in male and female hybrid embryos. (C) A schematic karyotype of a hybrid nucleus with sites of FISH probe hybridization highlighted. (D) Interphase nuclei from brain cells of male and female larvae have wild type organization of the dodeca and 2L3L satellites (see Figure 3.6 for wild type control). Hybrid larvae were generated a cross between *D. melanogaster yv* females and *D. simulans v* males, and were sexed using mouth hook coloration.

3.3.6 *cis*-by-*trans* regulatory divergence causes functional divergence at *D. melanogaster* and *D. simulans Lhr*.

To address whether *Lhr* regulation has diverged between *D. melanogaster* and *D. simulans*, I surveyed *Lhr* transcript levels in three strains from each of the two species, and found no significant difference (Figure 3.10A). I also found that the expression of ϕ\{Dmel\Lhr-HA\} and P\{Dsim\Lhr-HA\} constructs was at wild type levels, in their own species background. However, qRT-PCR showed that the level of *Lhr* transcript from ϕ\{Dsim\Lhr-HA\} transgene in *D. melanogaster* was significantly higher in comparison to the ϕ\{Dmel\Lhr-HA\} transgene, which is indicative of *cis*-regulatory divergence (Figure 3.10B). This upregulation is clearly a function of the *D. melanogaster* background, because transcription of the same construct in *D. simulans* is at wild type levels. This pattern suggests that *Lhr* has undergone *cis*-by-*trans* compensatory regulation; i.e. *cis*-regulatory regions and *trans*-factors have co-evolved within each species to maintain a constant level of gene expression, and the uncoupling of such species-specific compensatory changes in a foreign genetic background is causing mis-expression [181]. I hypothesized that such a mechanism might cause asymmetric expression of *Lhr* orthologs in hybrids and by extension underlie the asymmetric rescue properties of *Lhr* orthologs.

In other words, is the *D. simulans* mutation rescuing hybrid sons because it removes a greater fraction of the total *Lhr* gene product, compared to a mutation in the *D. melanogaster* ortholog? To test this hypothesis, I did allele-specific pyrosequencing to estimate the relative expression levels of the two *Lhr* orthologs in hybrids. I examined 2-3 day old larvae because temperature shift experiments have shown that the L2/L3 stage is the critical phase of the lethality [50].

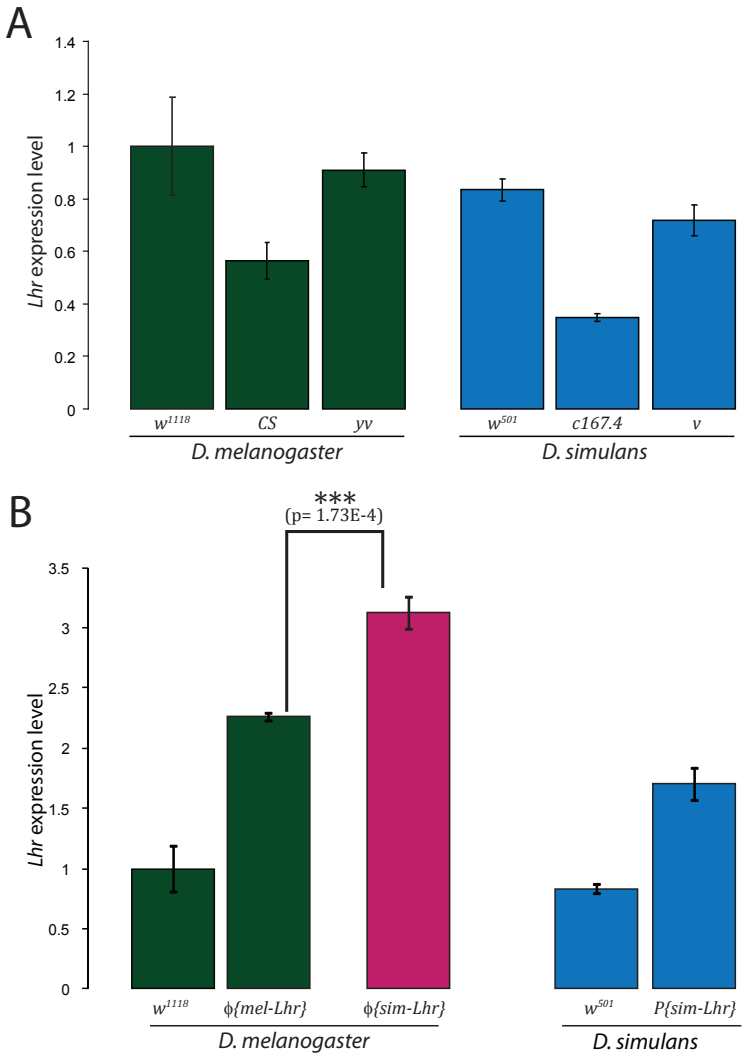

Figure 3.10: *Cis*-by-*trans* regulatory divergence of *D. simulans Lhr* (A) *Lhr* transcript levels in different wild type and marker strains of *D. melanogaster* and *D. simulans* (Nested ANOVA $F_{1,4} = 0.89$; $p = 0.39$). (B) *Lhr* transcript levels in transgenic lines compared to the corresponding host strain genetic background (w^{1118} for *D. melanogaster* and w^{501} for *D. simulans*). The transgenic lines are homozygous for the transgene and for the endogenous *Lhr* allele, and therefore have four copies of *Lhr*. *Lhr* transcript abundance is doubled in $\phi\{mel\text{-}Lhr\}$ and $P\{sim\text{-}Lhr\}$ relative to their respective reference backgrounds, indicating wild type expression levels of these transgenes in their native species. In contrast, *Lhr* transcription in $\phi\{sim\text{-}Lhr\}$ is significantly higher than in $\phi\{mel\text{-}Lhr\}$ (by two-tailed t-test) and is ≈3X the level of the reference background. For both A and B, RNA was isolated from 6-10 hr old embryos. *Lhr* expression levels were measured relative to *rpl32* using quantitative RT-PCR. Expression levels were normalized by setting *D. melanogaster* w^{1118} strain to 1. Error bars represent standard deviation within biological replicates, n≥6 for all except $P\{sim\text{-}Lhr\}$ where n=3.

As expected *Lhr* transcript from the pure species parents was 100% for their respective species-specific SNP. However, there was a significant overrepresentation of the *D. simulans*-specific SNP in both hybrid males and females, with ≈65% of *Lhr* deriving from the *D. simulans* ortholog in hybrid males and ≈60% in hybrid females (Figure 3.11). These data confirm the expectation that *cis*-by-*trans* divergence of *Lhr* regulation causes asymmetric expression in hybrids.

3.4 Discussion

3.4.1 Function underlying hybrid lethal activity of *Lhr* was present in the ancestral allele.

Lhr and *Hmr* are D-M interaction partners that cause hybrid lethality [51]. Population genetic analysis on *Lhr*, *Hmr* and other HI genes has found them to be evolving rapidly under selection [51, 127, 112, 88, 90, 8, 97]. The key assumption often made is that this selection-driven divergence is the molecular basis of incompatibility in hybrids. A direct expectation from this assumption is that independently evolving orthologs of a D-M gene should be non-equivalent with respect to the HI phenotype. Genetic data support this expectation for *Lhr*, because a loss of function mutation in *D. simulans Lhr* rescues lethal hybrid sons, while a loss-of-function mutation in *D. melanogaster Lhr* does not [51]. These findings led to the hypothesis that HI is specifically due to divergence specific to the *simulans* lineage, and is caused by a change in its heterochromatin association properties.

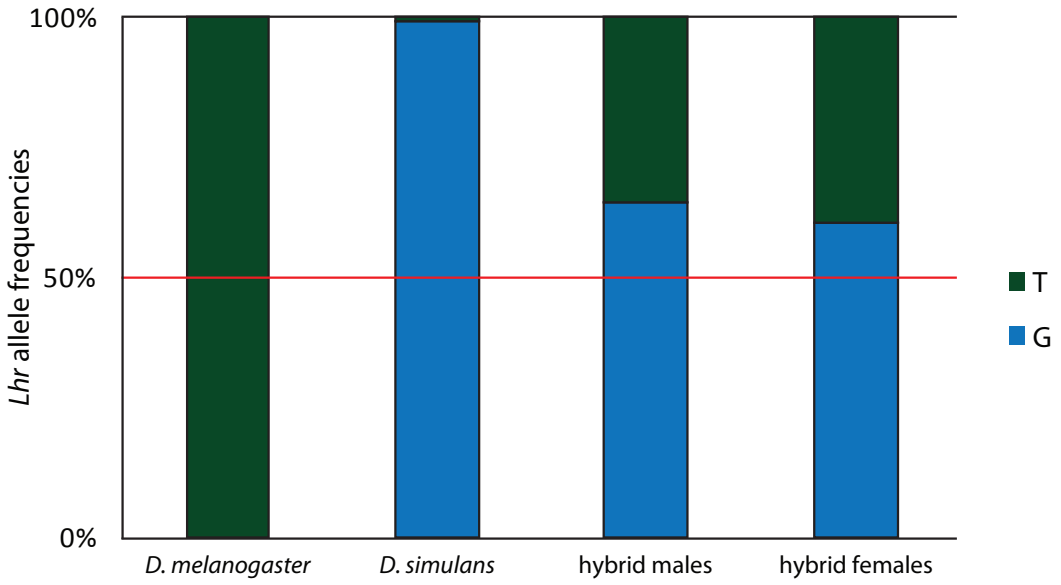

Figure 3.11: Asymmetric expression of *Lhr* orthologs in hybrids. Pyrosequencing across a SNP fixed between *Lhr* orthologs was used to measure the ratio of allelic transcription in pure-species and hybrid larvae that were 3-5 days old. *Lhr* transcript from each of the pure species is 100% for the species-specific variant of the SNP, while the *D. simulans*-specific SNP is detected at levels significantly greater than the expected 50% in both male and female hybrids ($p < .0001$ and $p = .005$ for hybrid males and females respectively, by two-tailed t-test)

Contrary to this expectation, Brideau et. al. (2011) [86] found that both *D. melanogaster* and *D. simulans Lhr* could cause HI and were unable to detect any differences in their heterochromatic properties. However, the use of GAL4-UAS driven *Lhr* transgenes as an assay limited their abilities to detect functional differences based on differences in expression level. I therefore, decided to use a native-promoter driven transgenic system to assay the functional divergence of *Lhr* orthologs. Consistent with Brideau et. al. (2011) I find that both *Lhr* orthologs suppress hybrid rescue by *D. simulans Lhr*[1], supporting the inference that hybrid lethal activity is a shared ancestral function. However, using a more sensitive interaction assay with *Hmr*, I did in fact detect that the lethal interaction was greater with *simulans Lhr* (Table 3.5), consistent with the pattern of genetic asymmetry where a mutation in *D. simulans Lhr* rescues hybrid lethality, while a deficiency removing *D. melanogaster Lhr* does not.

The use of native promoters allowed us to exclude the possibility that the shared incompatibility is an artifact of over-expression; a particular concern in the hybrid background, as it is less likely to be buffered against the adverse effects of excess chromatin proteins. I also used more upstream DNA than Prigent et. al. (2009) [182], whose transgene was also functional. Despite this, the transgenic system appears to have greater hybrid lethal activity than the endogenous locus in genetic tests (Table 3.4). However, this anomalous behavior of the transgene does not undermine the conclusion that the *D. melanogaster Lhr* ortholog has hybrid lethal activity; instead it highlights the finding that the hybrid male background is extremely sensitive to the dosage of *Lhr*, revealing a functional divergence that, at least in part, depends on expression levels.

3.4.2 Conserved heterochromatic properties of *Lhr* orthologs.

LHR orthologs in multiple Drosophila species associate with heterochromatin, at least in a *D. melanogaster* background [86]. However, the heterochromatic landscape is dramatically different in each species [11]. This raises the possibility that rapid evolution of *Lhr* orthologs could be reflecting functional divergence necessitated by association with fast-evolving heterochromatic sequences.

I addressed this possibility for LHR in two ways, (1) I mapped its localization within *D. melanogaster* pericentric heterochromatin and (2) compared localization patters of LHR in *D. simulans*, and (3) examined localization of sim-LHR in *D. melanogaster* background. In both species LHR localized to DAPI-rich heterochromatic foci. This localization to heterochromatin was however, noticeably not ubiquitous. For example, LHR signal showed no overlap with the AATAT satellite or the 359 bp satellite-block, which together account for a major fraction of *D. melanogaster* pericentric repeats [176]. In contrast, LHR consistently co-localized with the dodeca satellite in both species during interphase. The conservation of this co-localization pattern was particularly striking, given the dramatic redistribution of this satellite between the two species. In *D. melanogaster* dodeca repeats are located within pericentric heterochromatin of the third chromosome, while in *D. simulans* dodeca has expanded to both autosomes, with the major focus on the second chromosome (see Figure 3.6A and [178]). It is clear therefore that the chromosomal distribution of LHR between the two species is divergent. However, despite this apparent divergence, *simulans* LHR when expressed in *D. melanogaster* co-localized perfectly with its endogenous ortholog (Figure 3.7), demonstrating full conservation of their heterochromatic localization properties.

Could the shared hybrid lethal activity of *Lhr* orthologs be a function of their conserved heterochromatic localization?

3.4.3 No evidence for heterochromatic defects in incompatible hybrids.

I addressed this question by analyzing localization patterns of LHR and the distribution of other heterochromatic markers, including the dodeca satellite, in hybrids. I found no evidence for heterochromatic defects in hybrids (Figure 3.9). Dodeca organization is unaffected in lethal males and viable females. Like in pure species, LHR localizes to heterochromatic foci, especially dodeca, in both hybrid males and female embryos. Although the cytological analyses done here argue against gross defects in heterochromatin, their resolution would not detect fine-scale perturbations. Global studies from the van Steensel group have also found hundreds of LHR localization sites within euchromatin [114], and it is possible that LHR is mislocalizing at euchromatic loci in hybrids. Nonetheless, these results set *Lhr* apart from the other two heterochromatin-associated HI genes, *OdsH* and *Zhr*, where incompatible hybrids have gross defects in heterochromatin [85, 111]. Such defects have been interpreted as support for the hypothesis that internal conflict with selfish-heterochromatic elements is driving HI [165]; by this logic these results then do not support this view for *Lhr* evolution.

3.4.4 Regulatory divergence underlies asymmetric hybrid lethal effects of *Lhr* othologs.

I therefore examined whether functional divergence of *Lhr* orthologs was reflecting regulatory sequence divergence rather than CDS divergence. Asymmetric expression of *Lhr* orthologs in the hybrid background could explain the aforementioned genetic asymmetry. I tested this hypothesis by measuring allele-specific expression of *Lhr* orthologs in hybrid larvae. The results strongly support this hypothesis; I found that 66% of the total *Lhr* transcripts in lethal hybrid male larvae originates from the *simulans* allele (Figure 3.11). Thus a mutation in *D. simulans Lhr* creates hybrid sons with only $1/3^{rd}$ the wild type level of Lhr transcript, while hybrid sons with a mutation in *D. melanogatser* ortholog have twice that amount. This pattern is consistent with a loss-of-function mutation in *D. simulans Lhr* producing viable hybrids because it removes a greater fraction of the total *Lhr* gene product, while a deficiency removing *D. melanogaster Lhr* produces inviable sons.

This data strongly argues that the hybrid male genotype has evolved an acute sensitivity to *Lhr* dosage. Taken together with the genetic experiments showing that hybrid lethal activity is a shared ancestral function, this provides compelling evidence to reinterpret the functional asymmetry in terms of regulatory divergence rather than CDS divergence.

This divergence, however, is not reflecting a species-specific adaptation in expression levels, because *Lhr* expression surveyed in multiple strains from the two species was not significantly different from each other (Figure 3.10A). Using these transgenes, I was able to locate the molecular basis of this dysregulation

in hybrids to the uncoupling of species-specific compensatory changes between *cis*-regulatory sequence and *trans*-factors. Interestingly, studies comparing the evolution of transcriptional networks between species have found this type of regulatory divergence most frequently associated with mis-expression in interspecific hybrids [181, 62]. *Cis*-by-*trans* regulatory divergence has been shown to cause a large number of relatively small changes in interspecific hybrids; this is the first time that it is implicated as the source of functional divergence at a D-M locus. This result underscores the importance of factoring in regulatory divergence in the interpretation of interspecies experiments.

3.4.5 A Dobzhansky-Muller interaction between a derived and an ancestral allele.

Results presented here demonstrate that hybrid lethal activity is a function shared by both *Lhr* orthologs but that there is in fact a quantitative difference between them in their potency to cause lethality. The interpretation that the molecular interaction causing HI was present in the ancestral allele of *Lhr* is further supported by the observation that GAL4-UAS driven expression of *D. yakuba Lhr*, an outgroup species, also kills hybrid sons [86]. Unlike *Lhr*, complementation tests with its D-M partner, *Hmr* showed that only the *melanogaster* ortholog but not the *D simulans* ortholog is capable of HI [127]. Thus implying that, the HI effect of *Hmr* was derived during its independent evolution in the *D. melanogaster* lineage. I propose two models that explain how incompatibility between the ancestral *Lhr* allele and a derived *Hmr* allele could evolve.

The two-locus model is a variation of the derived-ancestral incompatibility

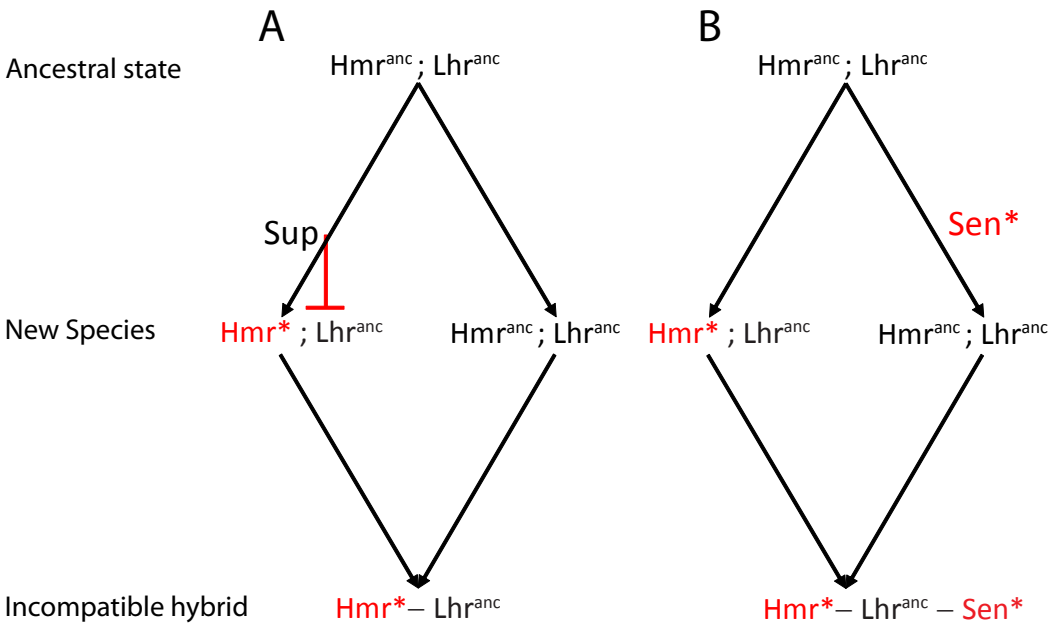

Figure 3.12: Alternative models for the evolution of hybrid lethality: incompatibility between an evolutionarily derived *D. melanogaster Hmr* and ancestral *Lhr*. In the first model a suppressor (Sup) fixed in the *melanogaster* lineage preventing the two-locus D-M interaction in the pure species background. In the second model an additional locus (Sen*) from the *simulans* lineage is needed to complete the lethal interaction.

described by Muller in his 1942 treatise [19]. In this model, a suppressor evolves first and is fixed in the *melanogaster* lineage, before the incompatibility-causing substitutions evolved in *Hmr* (Figure 3.12A). In the hybrid background, the suppressor is diluted, exposing the lethal interaction. Alternatively incompatibility could result from a complex epistatic interaction involving three or more loci. Here, evolution along the *simulans* lineage creates a sensitizing locus, *Sen**, that causes the hybrid background to become sensitive to the dosage of *Lhr* in the presence of *Hmr* from the *melanogaster* lineage (Figure 3.12B). I favor the three-locus model because in the two-locus model over-expression of *Lhr* orthologs in *D. melanogaster* is expected to at least partially overcome the suppressor and create the incompatible interaction. However, GAL4-UAS over-expression of either *Lhr* ortholog has no effect in a *D. melanogaster* pure species background [51, 86].

Other examples of ancestral-derived incompatibilities have also been discovered, such as the inter-allelic incompatibility at the *S5* locus in rice, and the bi-allelic incompatibility between the derived *S. cerevisiae* splicing factor *MRS1* and the ancestral *COX1* mRNA [76, 78]. However, unlike the incompatible *S5* alleles which differ by only two amino acid substitutions, and *MRS1* where the incompatibility has been mapped to three amino acid substitutions, *Lhr* orthologs have diverged rapidly under selection. It is therefore remarkable that despite striking protein sequence divergence between the hybridizing species, hybrid lethality has evolved as sensitivity to the dosage of an ancestral function. The key mechanistic insight from this is that instead of looking for a process or cellular component that differentiates *Lhr* orthologs as the source of hybrid lethality, I now know that the senstivity to *Lhr* in hybrids is based on a function and/or interaction that is common to both orthologs.

3.4.6 Role of selection in the evolution of incompatibilities.

There are least 6 HI genes known that are rapidly diverging under selection. With the exception of *OdsH* and *Prdm9*, where the signature of selection is restricted to a single functional domain [85, 97], in all other cases peaks of non-synonymous substitutions do not coincide with a particular domain within the HI gene. In these cases, it has been assumed that changes derived under selection have led to functional divergence, in turn causing incompatibility. However, it remains to be tested if that is truly the case.

Here I have tested the hybrid lethal activity of both *Lhr* orthologs and have found that despite striking divergence of the CDS, hybrid lethal activity is a shared ancestral function. I have not ruled out the contribution of CDS divergence to differences in hybrid lethal activities. However, it appears that the D-M asymmetry at the *Lhr* locus is largely explained by regulatory divergence. These results highlight the complexity of the interspecific background and emphasize that hybrids are far from being the stoichiometric sum of two parental genomes. I suggest that although positive selection in the CDS is a characteristic of HI genes, the phenotypic target of selection and its consequence to HI might be much less direct than expected.

CHAPTER 4

AN ANCESTRAL INDEL POLYMORPHISM CONTRIBUTES TO THE FUNCTIONAL DIVERGENCE OF *LHR* ORTHOLOGS

4.1 Introduction

What is the evolutionary driver of speciation? A significant step towards answering this question has been the identification of hybrid incompatibility (HI) genes, i.e. genes with "incompatible substitutions" that cause breakdown in interspecific hybrids. The next challenge is describing the evolutionary basis for the origin of such "incompatible substitutions". The classic Dobzhansky-Muller (D-M) model elegantly explains how substitutions incompatible only in an interspecific context can evolve, however it is agnostic on the nature of the intraspecific evolutionary forces that cause them (reviewed in [183, 184]). The model is equally consistent with incompatible substitutions evolving as functionally neutral mutations drifting to fixation or as functionally advantageous mutations being driven to fixation by natural selection.

It is therefore particularly intriguing that so many HI genes show high rates of sequence divergence driven by positive selection. If this divergence corresponds to the "incompatible substitutions" then there is a direct link between the phenotype under selection and HI. This is very likely for the hybrid sterility gene *OdsH*, where the signature of selection is concentrated within the DNA-binding homeodomain [8]. Functional analysis of *OdsH* orthologs has implicated divergent DNA-binding activity in hybrid incompatibility [85]. However, such a direct link between sequence divergence and function remains to be established for other rapidly evolving HI genes.

The HI gene *Lhr* poses an interesting paradox. Like many other HI genes, *Lhr* orthologs have strikingly divergent sequences evolving under positive selection [51]. The classic D-M model describes HI as the negative ectopic interaction between two derived loci, thus setting up the expectation that selection-driven divergence of *Lhr* created "incompatible substitutions" in one of the hybridizing lineages. Surprisingly, however in transgenic assays *Lhr* orthologs from both species can cause hybrid dysfunction [86]. This argues against the expectation that the hybrid lethal activity of *Lhr* is solely the outcome of selection-driven substitutions in its protein coding sequence (CDS) specific to *D. simulans*. Moreover, results presented in chapter 3 argue that the divergent hybrid lethal activities of *Lhr* orthologs can be largely attributed to their asymmetric expression in the hybrid background. The *D. simulans Lhr* allele is expressed two-fold higher than the *melanogaster* ortholog in the F1 hybrid. But it is still an open question whether divergence of the CDS might also be contributing to the differential hybrid lethal effects of *Lhr*.

Lhr orthologs have ≈50 fixed differences between *D. melanogaster* and *D. simulans* scattered throughout a protein sequence of ≈330 residues. Additionally, *Lhr* from each of the sibling species *D. simulans*, *D. mauritiana* and *D. sechellia* has a 16 amino acid insertion relative to the *D. melanogaster* ortholog. The insertion is absent in the outgroup species and is therefore identified as a derived state, specific to the common ancestor of the sibling species. This 16 amino acid insertion is also interesting because it may affect the structure of a predicted leucine zipper in the LHR protein, and had been proposed as a candidate for the functional divergence of *Lhr* orthologs [51].

The discovery of *D. simulans* Lhr^2, a hybrid rescue line segregating an un-

usual *Lhr* allele missing this 16 amino acid insertion, motivated us to further explore the effect, if any, of this indel on the hybrid lethal activity of *Lhr* orthologs. However, the Lhr^2 alelle also has several rare polymorphisms and a complex deletion in its C-terminus within a sequence of high conservation (Figure 4.1). Thus, even if the hybrid rescue property of the *D. simulans* Lhr^2 strain is a function of the unusual CDS of the Lhr^2 allele, it is unclear whether one of these changes is individually responsible or if it is the combined effect of multiple sequence variants.

A population survey revealed that the ancestral non-insertion form is segregating at a very low frequency in some *D. simulans* populations [185]. Nolte et. al. (2008) tested 2 *D. simulans* strains that carried *Lhr* alleles lacking the 16 amino acid insertion but not the C-terminal deletion in hybrid crosses. Neither of these strains produced viable hybrid sons, leading them to conclude that the hybrid rescue property of the *D. simulans* Lhr^2 strain could not be attributed to the ancestral non-insertion form of the 16 amino acid indel, leaving the complex C-terminal mutation as the most likely candidate. However, whether the indel polymorphism contributes to divergent hybrid lethal activities of *Lhr* orthologs remained untested.

Although indels are a common type of sequence variation, they are rarely considered in evolutionary studies. The reason for this is that their origins and functional consequences are poorly understood. Analysis of indels within protein sequences supports the view that they effect protein-folding, and computational analysis of high-throughput protein interaction data sets suggests that indels modify protein interaction interfaces, thereby significantly rewiring the interaction networks [186, 187]. Moreover, studies comparing patterns of evolu-

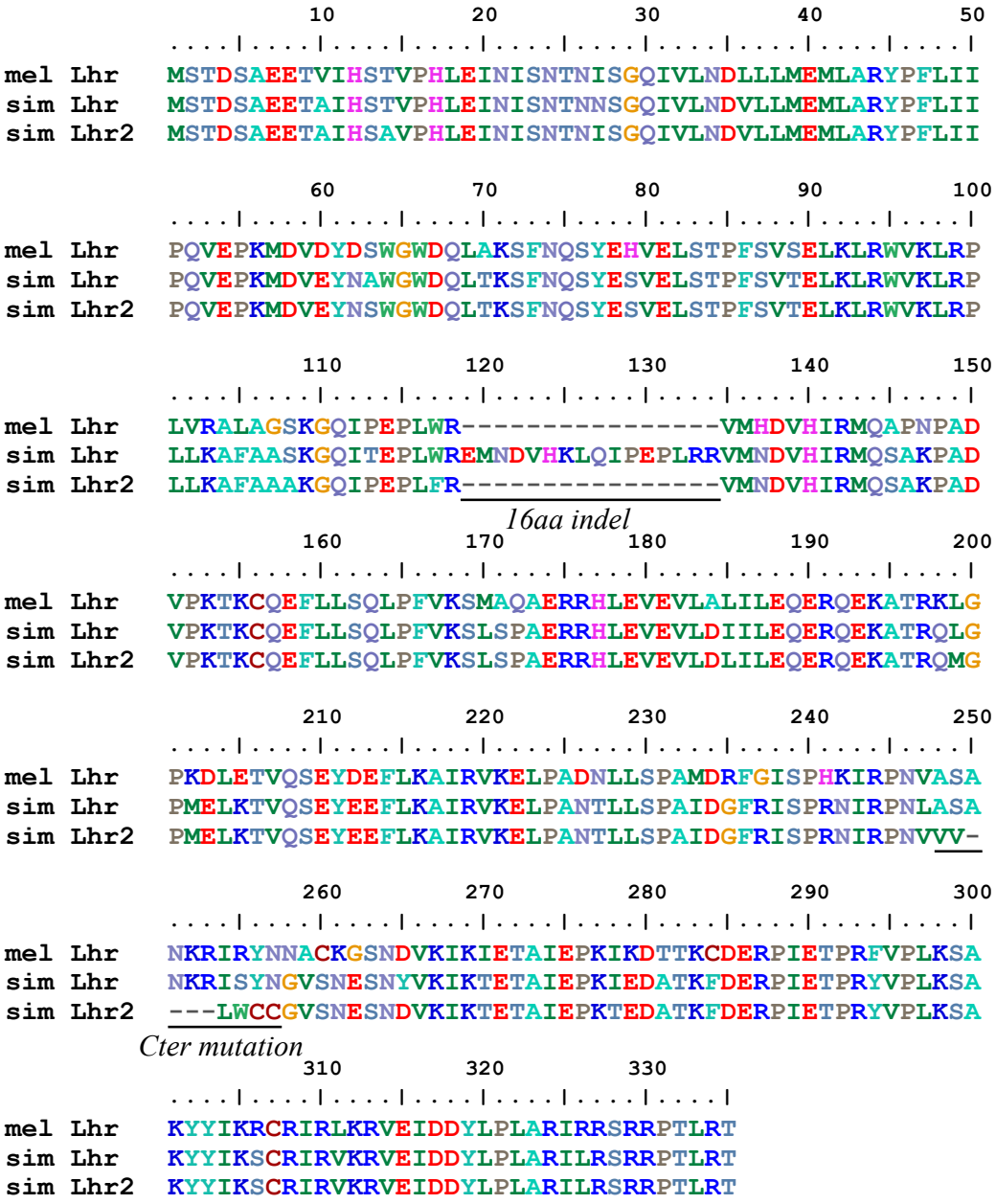

Figure 4.1: Alignment of *D. simulans Lhr*^{2} protein sequence with wild type orthologs. The 16 amino acid indel polymorphism and C-terminal mutation are underlined.

tion of *Catsper1*, a sperm-specific calcium channel, found evidence for positive selection for elevated rates of indel substitutions within its intracellular domain across multiple primate and rodent species [188, 189]. The authors suggest that the selection for indels might be a consequence of their effect on the regulation of the *Catsper1* channel, which can affect sperm motility, an important determinant in sperm competition.

In this study I first confirm that the hybrid rescue property of the *D. simulans* Lhr^2 strain is indeed a function of the CDS of the Lhr^2 allele. I then use the unusual Lhr^2 allele to address the key question of whether segregating ancestral polymorphisms are simply neutral markers of the speciation process, or instead have functional consequences.

4.2 Materials and Methods

4.2.1 DNA Constructs

To generate constructs for transgenic experiments, first the wild type *Lhr* CDS in p{sim-Lhr} was replaced by the Lhr^2 CDS using a three-piece fusion PCR strategy. The first and last PCR products, containing upstream and downstream genomic regions, were amplified using p{sim-Lhr} as the template, with primer pairs 691/938 and 941/664, respectively. The central PCR product containing the Lhr^2 CDS was amplified from *D. simulans* Lhr^2 genomic DNA, with primer pair 939/940. The three overlapping PCR products were then used as templates for the fusion PCR, cloned into the pCR-BluntII vector to create the plasmid p{sim-Lhr2}, and sequenced completely. A triple-HA tag in-frame with the C-

terminus of *Lhr*^{2} CDS was synthesized using a two-piece fusion PCR strategy. Two overlapping PCR products were amplified using p{sim-Lhr2} as the template, with primer pairs 882/728 and 729/664. This fusion PCR product was TOPO cloned into the pCR-BluntII vector. This intermediate construct was digested with *Sac*II and *Apa*I and the fragment released was subcloned into p{sim-Lhr2}, generating p{sim-Lhr2-HA}. The full insert was sequenced completely and subcloned into the MCS of pCasper4*attB* using *Not*I and *Kpn*I restriction enzymes.

To synthesize the construct p{sim-Lhr2-HA + 16aa}, the 16 amino acid insertion was inserted into the *Lhr*^{2} CDS using a two-piece fusion PCR strategy. The two overlapping PCR products were amplified using p{sim-Lhr2-HA} as the template, with primer pairs 691/945 and 946/664. These fragments were used as templates for the fusion PCR, and the gel-purified product was TOPO cloned into the pCR-BluntII vector and sequenced completely. The insert was then subcloned into pCasper4*attB* exactly as in p{sim-Lhr2-HA}. The construction of p{sim-Lhr2-HA +Cter} where the complex mutation in the C-terminus mutation in *Lhr*^{2} CDS was replaced by 10 residues of wild type *D. simulans Lhr* sequence, was done as above using primer pairs 691/942 and 943/664.

For yeast two-hybrid experiments the *Lhr*^{2} CDS was amplified using primer pair 404/405 and cloned into pENTR-DTOPO (Invitrogen) according to the manufacturer's instructions, and verified by sequencing. The entry vector was recombined with the destination vectors in a standard LR Clonase-mediated reaction. The destination vectors used were pGADT7-AD and pGBKT7-DNA-BD (K. Ravi Ram, A. Garfinkel, and M.F. Wolfner, Cornell University; personal communication).

Table 4.1: Primers used in the materials and methods

No.	Sequence	Capitalized region
404	caccatgagtaccgacagcgccgaggaa	
405	tcatgttctcagcgtaggccg	
409	gtagctttctcttggcgctctt	
410	gtaagtgaactgaagctgcgttgg	
664	tcgcatAAGCTTctggcaggtggtaaccgatacgg	HindIII
691	tactatAAGCTTtggttgttccacacgactttatcg	HindIII
728	TGCATAGTCCGGGACGTCATAGGGATA GCCCGCATAGTCAGGAACATCGTATGG GTACATtgttctcagcgtaggccg	3xHA tag
729	CCCTATGACGTCCCGGACTATGCAGGA TCCTATCCATATGACGTTCCAGATTACG CTtgactttctttcgtataaaatgc	3xHA tag
882	tgtcgcccgcggaacgtcgcc	
938	cgtttcctcggcgctgtcggtactcat	
939	atgagtaccgacagcgccgaggaaacg	
940	tcatgttctcagcgtaggccgcctgg	
941	ccaggcggcctacgctgagaacatga	
942	ccaTTATAGCTTATTCTTTTATTGGCACTT Gctacgttgggtcttatgttgcg	C-ter fill-in
943	CAAGTGCCAATAAAAGAATAAGCTATA Atggtgttagcaatgaatcaaatgatgtc	C-ter fill-in
945	GATTTGCAATTTGTGTACATCGTTCATC TCCCGCCACAGAGGTTCAGTgattt gcccttggcagccgc	16aa fill-in
946	ACTGAACCTCTGTGGCGGGAGATGAA CGATGTACACAAATTGCAAATCcctgaacc tctgtttcgggtg	16aa fill-in

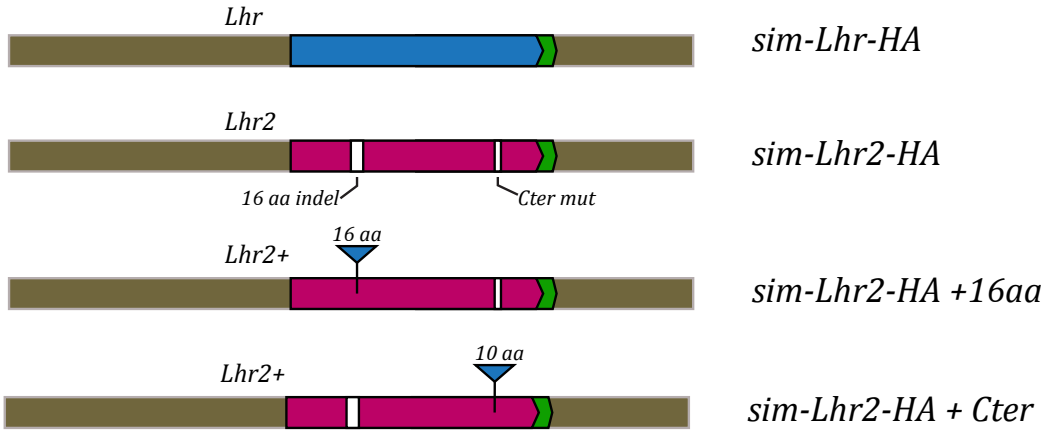

Figure 4.2: A schematic of *Lhr*^{2} constructs. All constructs contain the full *Lhr* coding sequences fused to the HA epitope tage (green) in the context of UTRs and genomic DNA from the *D. simulans* w^{501} strain. The white boxes represent the 16 amino acid indel and C-terminal mutation. Triangles represent replacement of *Lhr*^{2} CDS with sequence from wild type *D.simulans Lhr*.

4.2.2 Transgenic fly lines

ϕC31-mediated transformants of *D. melanogaster* were performed by Genetic Services. The integrations site used was M{3xP3-RFP.attP}ZH-86Fb [168] at cytological position 86Fb. Site-specificity of integrations were tested using the PCR assays described in chapter 3.

4.2.3 Recombination mapping of the *D. simulans Lhr*^{2} rescue activity

The *D. simulans Lhr*^{2} rescue strain was outcrossed to the non-rescuing *D. simulans v* strain. From this seven independent recombination lines were established by backcrossing 8-10 F1 daughters to 8-10 males from the *D. simulans v* strain. Sons from this cross were used to set up three hybrid crosses. Each hybrid cross was set up with approximately 30 recombinant sons, aged for 3 days, and 20 0-1 day old virgin *D. melanogaster* w^{1118} females. Viable F1 hybrid sons, which by definition inherit the rescue property, were PCR genotyped for their *Lhr* alleles. In order to determine if hybrid sons inherited the wild type *Lhr* or the *Lhr*^{2} allele from the *D. simulans* father, I used primer pairs 409/410 to PCR across the 16 amino acid indel. If sons inherit wild type *D. simulans Lhr* I expect to see two bands, the smaller band corresponding to the ancestral state in *D. melanogaster Lhr* and the larger size corresponding to the insertion in wild type *D. simulans Lhr*; however if they inherit the *Lhr*^{2} alelle, I expect to see only one band corresponding to the ancestral state.

4.2.4 RT-PCR, immunofluorescence and yeast two-hybrid

RT-PCR and immunofluorescence were performed as previously described in Chapter 3. Yeast two-hybrid assays were performed as in Brideau & Barbash (2011).

4.2.5 Phylogenetic analysis.

Phylogenetic tree was built by MEGA 3.1 using the maximum parsimony method [138]. The *Lhr* alleles used for the analysis are published in Brideau et. al. (2006) [51].

4.3 Results

A cross between wild type *D. melanogaster* females and *D. simulans* males produces only sterile daughters, and no sons. The genetic basis of male lethality appears to be fixed between the two species, as crosses between scores of different wild type strains fail to produce hybrid sons [128, 129]. The only two exceptions discovered are strains with mutations in *D. melanogaster Hmr* and *D. simulans Lhr* [130, 166].

Although previous analyses implicitly assumed that rescue in the *D. simulans* Lhr^2 strain is due to its unusual *Lhr* allele, this point has not been established. I therefore first did a crude mapping experiment to test whether the hybrid rescue function is associated with the Lhr^2 locus. I outcrossed *D. simulans* Lhr^2 to wild type *D. simulans* and tested for linkage between the Lhr^2 lo-

cus and hybrid rescue. I genotyped 48 viable hybrid sons, which by definition have inherited the rescue property, and found that all of them also inherited the Lhr^2 allele from the *D. simulans* parent. This pattern of co-segregation supports the hypothesis that the Lhr^2 allele is responsible for suppressing hybrid male lethality instead of an unrelated mutation segregating in the same genetic background.

I next sequenced 4 kb of genomic DNA spanning the Lhr^2 locus and found no insertions, deletions or rearrangements in its non-coding regions that might affect local gene regulation. Using quantitative RT-PCR I determined that *Lhr* expression in *D. simulans* Lhr^2 is not significantly different from wild type (t-test p = 0.2) (Figure 4.3A), demonstrating that the hybrid rescue property of *D. simulans* Lhr^2 is not due to a mutation affecting the expression level of *Lhr*. The Lhr^2 CDS is unusual in three respects (Figure 4.1). (1) Lhr^2 lacks the 16aa insertion, relative to *D. melanogaster Lhr*, that is present in frequencies near fixation in *Lhr* alleles from the sibling species. (2) Lhr^2 has a complex mutation in a conserved sequence near its C-terminus, which includes a 12 bp in-frame deletion and non-synonymous mutations causing unique substitutions in 6 adjacent amino acids. (3) 10 substitutions in the predicted Lhr^2 protein sequence are different from the consensus *D. simulans Lhr* [185].

While 5 of the 10 substitutions are rare, 5 are shared with *D. melanogaster Lhr*. With the *melanogaster*-like ancestral state at the 16 aa indel, it raised the possibility that Lhr^2 is a recent introgression of *D. melanogaster Lhr* into *D. simulans*. This was, however, rejected by phylogenetic analysis that firmly groups Lhr^2 with alleles from the sibling species (Figure 4.3B).

To test conclusively whether the Lhr^2 allele is defective for hybrid lethal ac-

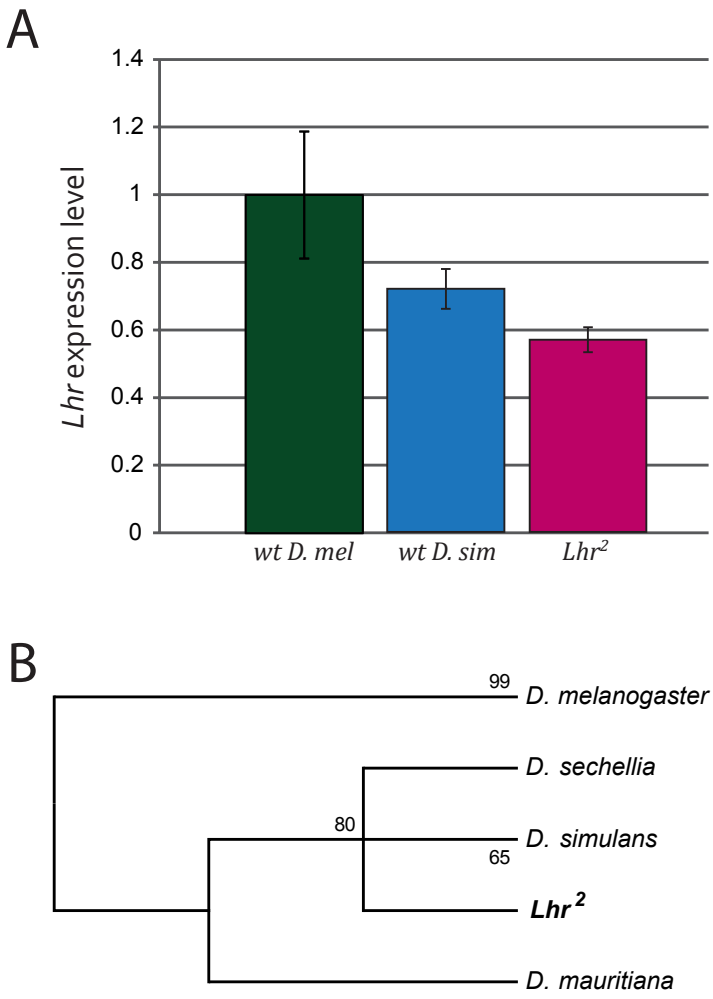

Figure 4.3: The Lhr^2 alelle is not an expression mutant or a *D. melanogaster* introgression. (A) Quantitative RT-PCR analysis comparing *Lhr* expression in *D. simulans* Lhr^2 with wild type *D. melanogaster* w^{1118} and *D. simulans v* strains. RNA was isolated from 6-10 hr old embryos. *Lhr* abundance was measured relative to *rpl32*. Expression levels were normalized by setting *D. melanogaster* w^{1118} strain to 1. Error bars represent standard deviation within biological replicates, n≥3 (B) The evolutionary history of Lhr^2 in the *melanogaster* subgroup was inferred using the Maximum Parsimony method. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) are shown next to the branches.

tivity, I set up a transgenic assay to compare it with wild type *D. simulans Lhr*. I used the ϕC31 site-specific integration system to generate a *D. melanogaster* strain carrying a *D. simulans* Lhr^2 transgene at the attP86Fb site on the third chromosome (Figure 4.2). The Lhr^2 CDS was C-terminally tagged with HA and placed under the control of wild type *D. simulans* regulatory sequences (from strain w^{501}), to generate the ϕ\{sim-Lhr2-HA\} construct.

Hybrid lethal activity was assayed using the *D. simulans* Lhr^1 complementation test established in chapter 3. *D. melanogaster* mothers heterozygous for one of the transgenes were crossed to *D.simulans* Lhr^1 fathers, Lhr^1 being a loss-of-function mutation that acts as a dominant suppressor of HI. If the transgene has hybrid lethal activity it is expected to suppress rescue by the Lhr^1 mutation. From experiments in chapter 3 I know that both *D. simulans* and *D. melanogaster* wild type *Lhr* orthologs fully suppress rescue by the *D. simulans* Lhr^1 mutation. In contrast, ϕ\{*sim-Lhr2-HA*\} only partially suppressed rescue, with viability in the range of 35-40% relative to the control class (Table 4.2 cross 2). This assay demonstrates that the Lhr^2 CDS has significantly reduced ability to cause HI but it is not a null allele. This conclusion is consistent with the observation that when crossed to the same *D. melanogaster* strain, *D. simulans* Lhr^1, which is an expression null, rescues hybrid male viability to 73%, while *D.simulans* Lhr^2 rescues at approximately 45% (data not shown).

Next, I wanted to determine the contribution of the two major structural mutations to the hybrid lethal activity of Lhr^2. To individually test the contribution of the complex C-terminal mutation and the 16aa deletion to hybrid lethal activity, each was individually replaced with wild type sequence in sim-Lhr2-HA to generate ϕ\{*sim-Lhr2-HA,+Cter*\} and ϕ\{*sim-Lhr2-HA,+16aa*\}, respectively

Table 4.2: Both structural mutations contribute to the reduced hybrid lethal activity of the *D. simulans Lhr*^{2} allele.

Cross	**Transgenic contruct**	**No. of hybrid females**	**No. of hybrid males**		**Relative viability of $\phi\{\}$(%)**
			Genotype 1 +/+	Genotype 2 $\phi\{\}$/+	
1	$\phi\{$*sim-Lhr*$\}$	135	74	0	0
2	$\phi\{$*sim-Lhr2*$\}$	494	226	80	35.4
		308	185	75	40.5
3	$\phi\{$*sim-Lhr2 +Cter*$\}$	269	175	0	0
		187	104	0	0
4	$\phi\{$*sim-Lhr2 +16aa*$\}$	337	178	28	15.7
		224	164	26	15.8

Crosses were between *D. melanogaster* females *w;*$\phi\{\}$/+ heterozygous for the different transgenes and *D. simulans Lhr*^{1} males. The transgenes carried a copy of the w^+ gene so the hybrid sons inheriting the transgene, $\phi\{\}$/+ (genotype 2) were distinguished from their +/+ siblings (genotype 1) by their eye-colour. All crosses were carried out at room temperature. Relative viability is the ratio of the number of hybrid sons inheriting the transgene (genotype 2) compared to the control class (genotype 1).

(Figure 4.2). Reverting the C-terminal mutation to the wild type sequence is sufficient for the $\phi\{$*sim-Lhr2-HA,+Cter*$\}$ transgene to fully complement rescue by *D. simulans Lhr*^{1} (Table 4.2 cross 3), demonstrating that this conserved region is necessary and sufficient for wild type hybrid lethal activity. Reverting the 16 amino acid deletion from the ancestral state to the derived insertion state also had a significant impact on the hybrid lethal activity of *Lhr*^{2}. The relative viability of hybrid sons inheriting $\phi\{$*sim-Lhr2-HA,+16aa*$\}$ was reduced to ≈15% (Table 4.2 cross 4), demonstrating that having the ancestral deletion state significantly contributes to the hybrid rescue activity of *Lhr*^{2}. Although I have only assayed the effect of this mutation on the background of the *Lhr*^{2} allele, it does suggest the possibility that this ancestral polymorphism is potentially contributing to the divergent hybrid lethal effects of *Lhr* orthologs.

LHR protein localizes to heterochromatin through interaction with Heterochromatin Protein1 (HP1) [51, 114]. I therefore asked whether the reduced hybrid lethal activity of *Lhr*^{2} was reflecting a defect in heterochromatin association. I performed yeast two-hybrid assays and found that the interaction between LHR protein encoded by the *Lhr*^{2} alelle and HP1 was indistinguishable from the wild type control (Figure 4.4A). Consistent with this result, LHR2-HA localized to heterochromatin *in vivo* and immuno-FISH experiments showed co-localization with the dodeca satellite, providing further support for wild type association with heterochromatin (Figure 4.4B). From these results I infer that the reduced hybrid lethal activity of *Lhr*^{2} is not because localization to heterochromatin is defective.

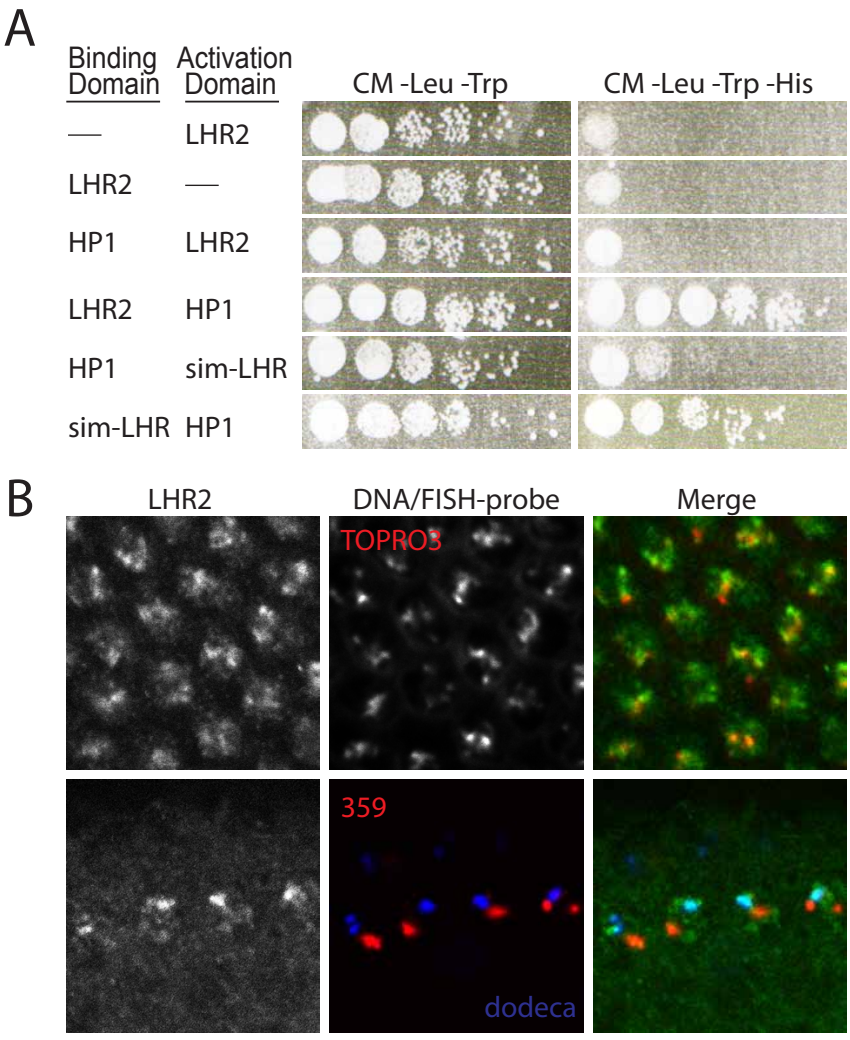

Figure 4.4: The sim-LHR2 protein interacts with HP1 and localizes to heterochromatin. (A) Interaction with HP1. Wild type *D. simulans* LHR was used as a positive control. Yeast two-hybrid interactions were detected by activation of HIS3 and growth on media lacking histidine; loading controls [complete media (CM) -Leu -Trp] contain histidine. (B) Localization of sim-LHR2-HA to heterochromatin in *D. melanogaster* cycle 12-14 embryos. Top, sim-LHR2-HA was detected with anti-HA (green) localizing to apical heterochromatin, detected by TOPRO3 staining of DNA at the embryo surface. Bottom, immuno-FISH experiment with anti-HA (green) detecting sim-LHR2-HA in interphase nuclei shows no overlap with the 359 bp (red) satellite but partially co-localizes with the dodeca satellite (blue).

4.4 Discussion

In this study I definitively identify the *Lhr*^{2} allele as a new hybrid rescue mutation; I then further show that this is due to changes in its CDS and use it to identify regions of the LHR protein critical for incompatibility. I devised modified *Lhr*^{2} alleles to individually assay specific regions for effects on hybrid lethal activity in *D. simulans Lhr*^{1} complementation tests. I find that a highly conserved stretch of 10 residues in the C-terminus of *Lhr* is critical for wild type levels of hybrid lethal activity. This is consistent with and supports results presented in chapter 3 as well as published results that *Lhr* orthologs from both species can cause incompatibility [86]. However, this region is clearly insufficient to completely explain the incompatible interaction, as more than half of the hybrid sons inheriting the ϕ{sim-Lhr2-HA} transgene die (Table 4.2 cross 2). It is likely that additional regions of the LHR protein also contribute to the incompatible interaction.

The effect of the 16aa indel polymorphism on hybrid lethal activity was excluded by Nolte et. al. (2008) using a population survey. They tested 2 *D. simulans* lines that retained the ancestral state of lacking the 16aa insertion, but neither of them rescued hybrid sons. However, in the transgenic assay I find a significant difference in hybrid lethal activity of the *Lhr*^{2} allele with and without the insertion (Table 4.2 cross 4). The lack of any phenotypic effects observed by Nolte et. al. (2008) is most likely because the effect of the 16aa indel is revealed only in a sensitized background. In this transgenic assay the C-terminal mutation in *Lhr*^{2} lowers the lethal activity of *Lhr*, providing us with the sensitivity to assess the contribution of the 16aa indel.

These results also speak to a focal question in the field of evo-devo, whether functional differentiation is the result of regulatory divergence or CDS divergence [190, 191]. I had previously shown that the different lethal activities of *Lhr* orthologs can be explained by divergent gene regulation (see Chapter 3). However, with this study, I can add the contribution of a CDS indel divergence to the species-specific differences in the strength of hybrid lethal activity between *D. simulans Lhr* and *D. melanogaster Lhr*.

Rigorous identification of incompatible substitutions had only been attempted for yeast. In case of *AEP2* it was narrowed down to multiple mutations within a region of 148 amino acids and in case of *MRS1* it was pared down to only 3 non-synonymous substitutions [43, 78]. There is no evidence of selection acting on either of these HI genes and both have experienced relatively limited sequence divergence. There are at least 6 HI genes known that are rapidly diverging under selection. Although it is implicitly assumed that this divergence is the basis of HI, this remains largely unexamined. Even for *Lhr*, where regions critical to incompatibility have now been identified, it is an open question whether any of the lineage-specific substitutions are relevant for HI.

Large structural polymorphisms are not unique to *Lhr*; other HI genes such as *Hmr* and *Prdm9* have multiple in-frame indels, as does the segregation distorter *RanGAP* [112, 97, 107]. So far the primary focus of evolutionary analysis has been single amino acid substitutions and indel variation has been largely ignored in the assessment of functional divergence. Recent high throughput analysis on human tissues has catalogued the occurrence of coding indels in hundreds of conserved and essential genes as well as in protein isoforms via alternative splicing, thus highlighting indels as an abundant source of structural

variation [192, 193, 194]. This characterization of the ancestral indel polymorphism in *Lhr* presents one functional argument supporting the prediction that coding indels play an important evolutionary role.

CHAPTER 5

CONCLUSIONS & FUTURE DIRECTIONS

Hybrid incompatibility (HI) genes are defined as loci that have a measurable effect on the fitness of hybrid progeny. *Lhr* and *Hmr* fit this criterion perfectly. A mutation in either *D. melanogaster Hmr* or *D. simulans Lhr* is sufficient to rescue lethal hybrid sons, generating a viable genotype instead. The protein coding sequence of both these genes has diverged dramatically between the hybridizing species, moreover population genetic analysis has found this divergence to be driven by natural selection. These observations suggested a simple and attractive hypothesis: functional differentiation favored by natural selection in one lineage would be the evolutionary basis of genetic incompatibility in interspecies offspring. By extension, the phenotypic target of selection represented by these loci would be, at least in part, culpable for the evolution of reproductive isolation between these species. Building a framework to test these hypothesis was the motivation behind this research. In order to articulate in greater detail the relationship between natural selection and HI the following questions can be asked: What is the function of *Lhr* in pure species? How has *Lhr* function diverged between the hybridizing species? What developmental process is breaking down in hybrids and does it correlate with the functional differentiation of *Lhr*? Can protein CDS substitutions, the outcome of natural selection, be directly implicated as the cause of HI?

When this research was begun the only known molecular function of LHR was its association with heterochromatic regions within the nucleus. Cytological analysis presented here has refined this association by examining its relationship to specific pericentric satellites within heterochromatin. Although it is now

clear that LHR has a specific distribution within heterochromatin, this study has not revealed the DNA or chromatin factors that configure this localization pattern. It has however, established that CDS divergence of *Lhr* orthologs is not causing different localization patterns. The functional differentiation evolved by selection remains obscure; further investigation into the biology of *Lhr* is needed to shed light on this issue.

Three major stumbling blocks to understanding the biology of *Lhr* have been: One, LHR has proven to be highly intractable to biochemical analysis. Two, in stark contrast with the hybrid situation, null mutation and over-expression genotypes of *Lhr* have no obvious phenotypes in a pure species background, thus making it intractable to standard genetic tests of function. Three, except for studies of HI *Lhr* lacks biological context. Besides association with heterochromatin, there is no compelling evidence in the public databases that would highlight a role for *Lhr* in existing protein interaction networks and/or developmental pathways. Hence the unbiased discovery of interaction partners through co-immunoprecipitation experiments is an obvious choice for the way forward. Given the transgenic tools developed in this thesis, such an experiment is now more doable. Another possible experiment similar to that done with HP1 can be designed, where LHR fused to the lac-repressor is targeted *in vivo* to a *lac*-operator array to determine if, and how, LHR affects chromatin and epigenetic states [195].

Because LHR associates with heterochromatin, the graveyard of selfish parasitic elements within the genome, it was suspected that the evolutionary origins of HI lie in conflict with selfish heterochromatic elements. The expectation was that this would be reflected in defects either in the localization of LHR or in the

structure and morphology of hybrid heterochromatin. However, I observed neither. Although the absence of such defects does not disprove the said hypothesis, it does provide a compelling argument for reconsidering the assumption that incompatibility in hybrids involves heterochromatin.

The ideal experiment for testing this relationship would be to uncouple LHR from heterochromatin and ask whether it can still cause incompatibility. Brideau N.J. designed such an experiment, he wanted to identify and then mutate the residues within LHR critical for HP1-interaction, and then assay this variant for hybrid lethal activity. However, this line of investigation had to be abandoned when technical issues confounded his efforts to whittle down the HP1-interaction domain. As the most simple and logical approach has failed I would like to suggest a couple of eccentric experiments to address this issue. Janssen et. al. (2000) [196] characterized specific polyamide compounds as tools to study chromatin structure. They showed that they target specific satellite repeats and that their binding mediated local opening of chromatin fibres. Remarkably, when fed to developing flies of a sensitized background, *white-mottled* and *brown-dominant*, they showed PEV suppression and homeotic transformations. It would be interesting to see if hybrid heterochromatin, and viability, was similarly responsive to the presence of such compounds. A second experiment involves altering the nuclear localization pattern of LHR by fusing it to domains with specific localization patterns, for example histone-recognition domains, and then assaying this variant for *D. simulans Lhr*^{1} complementation. If redirecting LHR to euchromatin mitigates its hybrid lethal activity that would strongly suggest a role for heterochromatin in HI.

Although research in this thesis has not revealed the developmental pathway that is malfunctioning in hybrids, it has revealed that this pathway is acutely sensitive to the dosage of *Lhr*. Several studies have suggested a link between dosage compensation and *Lhr* dependent hybrid male lethality [197, 198]. Interestingly, another HP1-interacting protein HOAP (also known as *caravaggio*) has recently been shown to have a role in regulating *Sxl*, the masterswitch of the sex-determination pathway [199]. Particularly they found a substantial zygotic component in the effect of HOAP on male viability; this is significant because *Lhr* also acts zygotically to cause incompatibility in hybrids. However, the critical phase of hybrid lethality is around second instar larval stage, significantly later than early embryogenesis when the critical dosage-sensitive decisions to establish *Sxl* are made. Additionally genetics has proven that lethality is specific to the *melanogaster* X chromosome and not the sex of the hybrid animal, arguing against incompatibility resulting from a simple misregulation of the sex-determination pathway. Thus, it is possible that hybrid lethality could be due to a partial failure of dosage compensation involving the melanogaster X chromosome. At present we know that the D-M interaction partner, *Hmr* is on the *D. melanogaster* X-chromosome, but it is not clear if it is the only factor responsible for the lethal effect of the *melanogaster* X. This could be genetically tested by crossing compound-X *D. melanogaster* females carrying a *mel-Hmr* transgene to wild type *D.simulans* males. If hybrid sons inheriting the *mel-Hmr* transgene are lethal, this would prove that *Hmr* is the only factor on the melanogaster X-chromosome required for HI.

The most striking result in this thesis is that hybrid lethal activity is not a derived function restricted to the *simulans* lineage, instead the observed differences in the hybrid lethal activities of *Lhr* orthologs rest in differences in their

cis-regulatory regions. Although it remains to be examined whether selection has played a role in evolving divergent regulation of *Lhr* orthologs, this result forces us to reconsider the role of CDS divergence, and consequently natural selection in the evolution of HI. To definitively test if CDS divergence makes any contribution to the different hybrid lethal effects of *Lhr* orthologs, *Hmr* interaction assays with transgenes containing *melanogaster* and *simulans Lhr* under the control of identical regulatory sequences need to be done.

Rapid evolution under natural selection is emerging as a signature of HI genes, and it is assumed that the incompatible interaction is the consequence of selection-driven substitutions. It is thus very intriguing that in the case of *Lhr* this appears to not be true. It requires that we consider more complicated scenarios, such as divergence of *Lhr* is reflecting functional differentiation of an essential developmental pathway. It is possible that while divergence of this developmental pathway is at the root of hybrid breakdown, CDS substitutions in *Lhr* are not directly causing hybrid breakdown. However, it is equally likely that selection driven evolution of *Lhr* in pure species is entirely unrelated to its function in hybrid breakdown. To resolve the relationship between selection and hybrid incompatibility a deeper understanding of the function of *Lhr* and the molecular mechanism underlying hybrid breakdown is needed.

BIBLIOGRAPHY

[1] Bomblies, K. Doomed lovers: mechanisms of isolation and incompatibility in plants. *Annu. Rev. Plant Biol.* **61**, 109–124 (2010).

[2] Brown, J. D. & O'Neill, R. J. Chromosomes, conflict, and epigenetics: chromosomal speciation revisited. *Annu. Rev. Genomics Hum. Genet.* **11**, 291–316 (2010).

[3] Presgraves, D. C. Darwin and the origin of interspecific genetic incompatibilities. *Am. Nat.* **176**, S45–60 (2010).

[4] Rieseberg, L. H. & Blackman, B. K. Speciation genes in plants. *Ann. Bot.* **106**, 439–455 (2010).

[5] Aruna, S., Flores, H. A. & Barbash, D. A. Reduced fertility of Drosophila melanogaster hybrid male rescue (Hmr) mutant females is partially complemented by Hmr orthologs from sibling species. *Genetics* **181**, 1437–1450 (2009).

[6] Sun, S., Ting, C. T. & Wu, C. I. The normal function of a speciation gene, Odysseus, and its hybrid sterility effect. *Science* **305**, 81–83 (2004).

[7] Barbash, D. A., Siino, D. F., Tarone, A. M. & Roote, J. A rapidly evolving MYB-related protein causes species isolation in Drosophila. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 5302–5307 (2003).

[8] Ting, C. T., Tsaur, S. C., Wu, M. L. & Wu, C. I. A rapidly evolving homeobox at the site of a hybrid sterility gene. *Science* **282**, 1501–1504 (1998).

[9] Naveira, H. & Fondevila, A. The evolutionary history of Drosophila buzzatii. XXI. cumulative action of multiple sterility factors on spermatogenesis in hybrids of D. buzzatii and D. koepferae. *Heredity* **67**, 57–72 (1991).

[10] Kamm, A., Galasso, I., Schmidt, T. & Heslop-Harrison, J. S. Analysis of a repetitive DNA family from Arabidopsis arenosa and relationships between Arabidopsis species. *Plant Mol. Biol.* **27**, 853–862 (1995).

[11] Lohe, A. & Roberts, P. chap. Evolution of satellite DNA sequences in Drosophila, 148–186. Heterochromatin: Molecular and Structural Aspects (Cambridge Univ. Press, Cambridge, 1988).

[12] Michalak, P. Epigenetic, transposon and small RNA determinants of hybrid dysfunctions. *Heredity* **102**, 45–50 (2009).

[13] Ferree, P. M. & Barbash, D. A. Distorted sex ratios: a window into RNAi-mediated silencing. *PLoS Biol.* **5**, e303 (2007).

[14] Ramsey, J. & Schemske, D. W. Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annu. Rev. Ecol. Syst.* **29**, 467–501 (1998).

[15] Navarro, A. & Barton, N. H. Accumulating postzygotic isolation genes in parapatry: a new twist on chromosomal speciation. *Evolution* **57**, 447–459 (2003).

[16] Noor, M. A., Grams, K. L., Bertucci, L. A. & Reiland, J. Chromosomal inversions and the reproductive isolation of species. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 12084–12088 (2001).

[17] Rieseberg, L. H. Chromosomal rearrangements and speciation. *Trends Ecol. Evol.* **16**, 351–358 (2001).

[18] Lynch, M. & Force, A. The origin of interspecific genomic incompatibility via gene duplication. *Am. Nat.* **156**, 590–605 (2000).

[19] Muller, H. J. vol. VI of *Biological Symposia*, chap. Isolating mechanisms, evolution and temperature, 71–125 (The Jaques cattell press, Lancaster Pennsylvania, 1942).

[20] Moyle, L. C., Muir, C. D., Han, M. V. & Hahn, M. W. The contribution of gene movement to the "two rules of speciation". *Evolution* **64**, 1541–1557 (2010).

[21] Masly, J. P., Jones, C. D., Noor, M. A., Locke, J. & Orr, H. A. Gene transposition as a cause of hybrid sterility in Drosophila. *Science* **313**, 1448–1450 (2006).

[22] Bikard, D. *et al.* Divergent evolution of duplicate genes leads to genetic incompatibilities within A. thaliana. *Science* **323**, 623–626 (2009).

[23] Mizuta, Y., Harushima, Y. & Kurata, N. Rice pollen hybrid incompatibility caused by reciprocal gene loss of duplicated genes. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 20417–20422 (2010).

[24] Yamagata, Y. *et al.* Mitochondrial gene in the nuclear genome induces reproductive barrier in rice. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 1494–1499 (2010).

[25] Hollocher, H. & Wu, C. I. The genetics of reproductive isolation in the Drosophila simulans clade: X vs. autosomal effects and male vs. female effects. *Genetics* **143**, 1243–1255 (1996).

[26] Masly, J. P. & Presgraves, D. C. High-resolution genome-wide dissection of the two rules of speciation in Drosophila. *PLoS Biol.* **5**, e243 (2007).

[27] Tao, Y., Chen, S., Hartl, D. L. & Laurie, C. C. Genetic dissection of hybrid incompatibilities between Drosophila simulans and D. mauritiana. I. differential accumulation of hybrid male sterility effects on the X and autosomes. *Genetics* **164**, 1383–1397 (2003).

[28] True, J. R., Weir, B. S. & Laurie, C. C. A genome-wide survey of hybrid incompatibility factors by the introgression of marked segments of Drosophila mauritiana chromosomes into Drosophila simulans. *Genetics* **142**, 819–837 (1996).

[29] Bhutkar, A., Russo, S. M., Smith, T. F. & Gelbart, W. M. Genome-scale analysis of positionally relocated genes. *Genome Res.* **17**, 1880–1887 (2007).

[30] Scannell, D. R., Byrne, K. P., Gordon, J. L., Wong, S. & Wolfe, K. H. Multiple rounds of speciation associated with reciprocal gene loss in polyploid yeasts. *Nature* **440**, 341–345 (2006).

[31] Kassahn, K. S., Dang, V. T., Wilkins, S. J., Perkins, A. C. & Ragan, M. A. Evolution of gene function and regulatory control after whole-genome duplication: comparative analyses in vertebrates. *Genome Res.* **19**, 1404–1418 (2009).

[32] Van de Peer, Y., Fawcett, J. A., Proost, S., Sterck, L. & Vandepoele, K. The flowering world: a tale of duplications. *Trends Plant Sci.* **14**, 680–688 (2009).

[33] Greig, D., Travisano, M., Louis, E. J. & Borts, R. H. A role for the mismatch repair system during incipient speciation in Saccharomyces. *J. Evol. Biol.* **16**, 429–437 (2003).

[34] Hunter, N., Chambers, S. R., Louis, E. J. & Borts, R. H. The mismatch re-

pair system contributes to meiotic sterility in an interspecific yeast hybrid. *EMBO J.* **15**, 1726–1733 (1996).

[35] Josefsson, C., Dilkes, B. & Comai, L. Parent-dependent loss of gene silencing during interspecies hybridization. *Curr. Biol.* **16**, 1322–1328 (2006).

[36] Dobzhansky, T. Studies on hybrid sterility. II. localization of sterility factors in Drosophila pseudoobscura hybrids. *Genetics* **21**, 113–135 (1936).

[37] Moehring, A. J., Llopart, A., Elwyn, S., Coyne, J. A. & Mackay, T. F. The genetic basis of postzygotic reproductive isolation between Drosophila santomea and D. yakuba due to hybrid male sterility. *Genetics* **173**, 225–233 (2006).

[38] Tao, Y., Zeng, Z. B., Li, J., Hartl, D. L. & Laurie, C. C. Genetic dissection of hybrid incompatibilities between Drosophila simulans and D. mauritiana. II. mapping hybrid male sterility loci on the third chromosome. *Genetics* **164**, 1399–1418 (2003).

[39] Harushima, Y., Nakagahra, M., Yano, M., Sasaki, T. & Kurata, N. Diverse variation of reproductive barriers in three intraspecific rice crosses. *Genetics* **160**, 313–322 (2002).

[40] Xu, Y., Zhu, L., Xiao, J., Huang, N. & McCouch, S. R. Chromosomal regions associated with segregation distortion of molecular markers in F2, backcross, doubled haploid, and recombinant inbred populations in rice (Oryza sativa L.). *Mol. Gen. Genet.* **253**, 535–545 (1997).

[41] Nakazato, T., Jung, M. K., Housworth, E. A., Rieseberg, L. H. & Gastony, G. J. A genomewide study of reproductive barriers between allopatric populations of a homosporous fern, Ceratopteris richardii. *Genetics* **177**, 1141–1150 (2007).

[42] Gadau, J., Jr, R. E. P. & Werren, J. H. Mapping of hybrid incompatibility loci in Nasonia. *Genetics* **153**, 1731–1741 (1999).

[43] Lee, H. Y. *et al.* Incompatibility of nuclear and mitochondrial genomes causes hybrid sterility between two yeast species. *Cell* **135**, 1065–1073 (2008).

[44] Gregorova, S. *et al.* Mouse consomic strains: exploiting genetic divergence

between Mus m. musculus and Mus m. domesticus subspecies. *Genome Res.* **18**, 509–515 (2008).

[45] Moyle, L. C. & Nakazato, T. Comparative genetics of hybrid incompatibility: sterility in two Solanum species crosses. *Genetics* **179**, 1437–1453 (2008).

[46] Hollingshead, L. A lethal factor in Crepis effective only in an interspecific hybrid. *Genetics* **15**, 114–140 (1930).

[47] Good, J. M., Handel, M. A. & Nachman, M. W. Asymmetry and polymorphism of hybrid male sterility during the early stages of speciation in house mice. *Evolution* **62**, 50–65 (2008).

[48] Mihola, O., Trachtulec, Z., Vlcek, C., Schimenti, J. C. & Forejt, J. A mouse speciation gene encodes a meiotic histone H3 methyltransferase. *Science* **323**, 373–375 (2009).

[49] Forejt, J. & Ivanyi, P. Genetic studies on male sterility of hybrids between laboratory and wild mice (Mus musculus L.). *Genet. Res.* **24**, 189–206 (1974).

[50] Barbash, D. A., Roote, J. & Ashburner, M. The Drosophila melanogaster hybrid male rescue gene causes inviability in male and female species hybrids. *Genetics* **154**, 1747–1771 (2000).

[51] Brideau, N. J. *et al.* Two Dobzhansky-Muller genes interact to cause hybrid lethality in Drosophila. *Science* **314**, 1292–1295 (2006).

[52] Orr, H. A. & Irving, S. Genetic analysis of the hybrid male rescue locus of Drosophila. *Genetics* **155**, 225–231 (2000).

[53] Sawamura, K., Yamamoto, M. T. & Watanabe, T. K. Hybrid lethal systems in the Drosophila melanogaster species complex. II. the Zygotic hybrid rescue (zhr) gene of D. melanogaster. *Genetics* **133**, 307–313 (1993).

[54] Coyne, J. A., Simeonidis, S. & Rooney, P. Relative paucity of genes causing inviability in hybrids between Drosophila melanogaster and D. simulans. *Genetics* **150**, 1091–1103 (1998).

[55] Presgraves, D. C. A fine-scale genetic analysis of hybrid incompatibilities in Drosophila. *Genetics* **163**, 955–972 (2003).

[56] Matute, D. R., Butler, I. A., Turissini, D. A. & Coyne, J. A. A test of the snowball theory for the rate of evolution of hybrid incompatibilities. *Science* **329**, 1518–1521 (2010).

[57] Sawamura, K. *et al.* Introgression of Drosophila simulans nuclear pore protein 160 in Drosophila melanogaster alone does not cause inviability but does cause female sterility. *Genetics* **186**, 669–676 (2010).

[58] Johnson, N. A. & Porter, A. H. Rapid speciation via parallel, directional selection on regulatory genetic pathways. *J. Theor. Biol.* **205**, 527–542 (2000).

[59] Auger, D. L. *et al.* Nonadditive gene expression in diploid and triploid hybrids of maize. *Genetics* **169**, 389–397 (2005).

[60] Graze, R. M., McIntyre, L. M., Main, B. J., Wayne, M. L. & Nuzhdin, S. V. Regulatory divergence in Drosophila melanogaster and D. simulans, a genomewide analysis of allele-specific expression. *Genetics* **183**, 547–61, 1SI–21SI (2009).

[61] Malone, J. H., Chrzanowski, T. H. & Michalak, P. Sterility and gene expression in hybrid males of Xenopus laevis and X. muelleri. *PloS One* **2**, e781 (2007).

[62] McManus, C. J. *et al.* Regulatory divergence in Drosophila revealed by mRNA-seq. *Genome Res.* **20**, 816–825 (2010).

[63] Michalak, P. & Noor, M. A. Genome-wide patterns of expression in Drosophila pure species and hybrid males. *Mol. Biol. Evol.* **20**, 1070–1076 (2003).

[64] Ranz, J. M., Namgyal, K., Gibson, G. & Hartl, D. L. Anomalies in the expression profile of interspecific hybrids of Drosophila melanogaster and Drosophila simulans. *Genome Res.* **14**, 373–379 (2004).

[65] Barbash, D. A. & Lorigan, J. G. Lethality in Drosophila melanogaster/Drosophila simulans species hybrids is not associated with substantial transcriptional misregulation. *J. Exp. Zool. B. Mol. Dev. Evol.* **308**, 74–84 (2007).

[66] Walia, H. *et al.* Dosage-dependent deregulation of an AGAMOUS-LIKE gene cluster contributes to interspecific incompatibility. *Curr. Biol.* **19**, 1128–1132 (2009).

[67] Michalak, P. & Noor, M. A. Association of misexpression with sterility in hybrids of Drosophila simulans and D. mauritiana. *J. Mol. Evol.* **59**, 277–282 (2004).

[68] Good, J. M., Giger, T., Dean, M. D. & Nachman, M. W. Widespread overexpression of the X chromosome in sterile F1 hybrid mice. *PLoS Genet.* **6**, e1001148 (2010).

[69] Lu, X. *et al.* Genome-wide misexpression of X-linked versus autosomal genes associated with hybrid male sterility. *Genome Res.* **20**, 1097–1102 (2010).

[70] Malone, J. H. & Michalak, P. Physiological sex predicts hybrid sterility regardless of genotype. *Science* **319**, 59 (2008).

[71] Rawson, P. D. & Burton, R. S. Functional coadaptation between cytochrome c and cytochrome c oxidase within allopatric populations of a marine copepod. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 12955–12958 (2002).

[72] Willett, C. S. Deleterious epistatic interactions between electron transport system protein-coding loci in the copepod Tigriopus californicus. *Genetics* **173**, 1465–1477 (2006).

[73] Dobzhansky, T. *Genetics and the origin of species* (Columbia University Press, New York, 1937).

[74] Muller, H. J. chap. Bearing of the Drosophila work on systematics., 185–268. The New Systematics (Oxford, The Clarendon Press, 1940).

[75] Orr, H. A. Dobzhansky, Bateson, and the genetics of speciation. *Genetics* **144**, 1331–1335 (1996).

[76] Chen, J. *et al.* A triallelic system of S5 is a major regulator of the reproductive barrier and compatibility of indica-japonica hybrids in rice. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 11436–11441 (2008).

[77] Long, Y. *et al.* Hybrid male sterility in rice controlled by interaction between divergent alleles of two adjacent genes. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 18871–18876 (2008).

[78] Chou, J. Y., Hung, Y. S., Lin, K. H., Lee, H. Y. & Leu, J. Y. Multiple molecu-

lar mechanisms cause reproductive isolation between three yeast species. *PLoS Biol.* **8**, e1000432 (2010).

[79] Christie, P. & MacNair, M. R. Complementary lethal factors in two North American populations of the yellow monkey flower. *J. Hered.* **75**, 510–511 (1984).

[80] Sweigart, A. L., Fishman, L. & Willis, J. H. A simple genetic incompatibility causes hybrid male sterility in mimulus. *Genetics* **172**, 2465–2479 (2006).

[81] Naveira, H. F. & Maside, X. R. chap. The genetics of hybrid male sterility in Drosophila, 330–338. Endless forms: species and speciation (Oxford University Press, Oxford, 1998).

[82] Davis, A. W. & Wu, C. I. The broom of the sorcerer's apprentice: the fine structure of a chromosomal region causing reproductive isolation between two sibling species of Drosophila. *Genetics* **143**, 1287–1298 (1996).

[83] Orr, H. A. & Irving, S. Complex epistasis and the genetic basis of hybrid sterility in the Drosophila pseudoobscura Bogota-USA hybridization. *Genetics* **158**, 1089–1100 (2001).

[84] Phadnis, N. & Orr, H. A. A single gene causes both male sterility and segregation distortion in Drosophila hybrids. *Science* **323**, 376–379 (2009).

[85] Bayes, J. J. & Malik, H. S. Altered heterochromatin binding by a hybrid sterility protein in Drosophila sibling species. *Science* **326**, 1538–1541 (2009).

[86] Brideau, N. J. & Barbash, D. A. Functional conservation of the Drosophila hybrid incompatibility gene Lhr. *BMC Evol. Biol.* **11**, 57 (2011).

[87] Yamamoto, E. *et al.* Gain of deleterious function causes an autoimmune response and Bateson-Dobzhansky-Muller incompatibility in rice. *Mol. Genet. Genomics* **283**, 305–315 (2010).

[88] Presgraves, D. C., Balagopalan, L., Abmayr, S. M. & Orr, H. A. Adaptive evolution drives divergence of a hybrid inviability gene between two species of Drosophila. *Nature* **423**, 715–719 (2003).

[89] Tang, S. & Presgraves, D. C. Evolution of the Drosophila nuclear pore

complex results in multiple hybrid incompatibilities. *Science* **323**, 779–782 (2009).

[90] Presgraves, D. C. & Stephan, W. Pervasive adaptive evolution among interactors of the Drosophila hybrid inviability gene, Nup96. *Mol. Biol. Evol.* **24**, 306–314 (2007).

[91] Barbash, D. A. Nup96-dependent hybrid lethality occurs in a subset of species from the simulans clade of Drosophila. *Genetics* **176**, 543–552 (2007).

[92] Sobel, J. M., Chen, G. F., Watt, L. R. & Schemske, D. W. The biology of speciation. *Evolution* **64**, 295–315 (2010).

[93] Schluter, D. Evidence for ecological speciation and its alternative. *Science* **323**, 737–741 (2009).

[94] Chou, J. Y. & Leu, J. Y. Speciation through cytonuclear incompatibility: insights from yeast and implications for higher eukaryotes. *BioEssays* **32**, 401–411 (2010).

[95] Boulton, A., Myers, R. S. & Redfield, R. J. The hotspot conversion paradox and the evolution of meiotic recombination. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 8058–8063 (1997).

[96] Myers, S. *et al.* Drive against hotspot motifs in primates implicates the PRDM9 gene in meiotic recombination. *Science* **327**, 876–879 (2010).

[97] Oliver, P. L. *et al.* Accelerated evolution of the Prdm9 speciation gene across diverse metazoan taxa. *PLoS Genet.* **5**, e1000753 (2009).

[98] Berg, I. L. *et al.* PRDM9 variation strongly influences recombination hotspot activity and meiotic instability in humans. *Nat. Genet.* **42**, 859–863 (2010).

[99] Charlesworth, B., Sniegowski, P. & Stephan, W. The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature* **371**, 215–220 (1994).

[100] Hurst, L. D., Atlan, A. & Bengtsson, B. O. Genetic conflicts. *Q. Rev. Biol.* **71**, 317–364 (1996).

[101] Brookfield, J. F. The ecology of the genome - mobile DNA elements and their hosts. *Nat. Rev. Genet.* **6**, 128–136 (2005).

[102] Wu, C. I., Lyttle, T. W., Wu, M. L. & Lin, G. F. Association between a satellite DNA sequence and the Responder of Segregation Distorter in D. melanogaster. *Cell* **54**, 179–189 (1988).

[103] Henikoff, S., Ahmad, K. & Malik, H. S. The centromere paradox: stable inheritance with rapidly evolving DNA. *Science* **293**, 1098–1102 (2001).

[104] Sandler, L. & Novitski, E. Meiotic drive as an evolutionary force. *Am. Nat.* **91**, 105–110 (1957).

[105] Zwick, M. E., Salstrom, J. L. & Langley, C. H. Genetic variation in rates of nondisjunction: association of two naturally occurring polymorphisms in the chromokinesin nod with increased rates of nondisjunction in Drosophila melanogaster. *Genetics* **152**, 1605–1614 (1999).

[106] Fishman, L. & Saunders, A. Centromere-associated female meiotic drive entails male fitness costs in monkeyflowers. *Science* **322**, 1559–1562 (2008).

[107] Presgraves, D. C. Does genetic conflict drive rapid molecular evolution of nuclear transport genes in Drosophila? *BioEssays* **29**, 386–391 (2007).

[108] Fossella, J. *et al.* An axonemal dynein at the Hybrid Sterility 6 locus: implications for t haplotype-specific male sterility and the evolution of species barriers. *Mamm. Genome* **11**, 8–15 (2000).

[109] Olds-Clarke, P. Models for male infertility: the t haplotypes. *Rev. Reprod.* **2**, 157–164 (1997).

[110] Tao, Y., Hartl, D. L. & Laurie, C. C. Sex-ratio segregation distortion associated with reproductive isolation in Drosophila. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 13183–13188 (2001).

[111] Ferree, P. M. & Barbash, D. A. Species-specific heterochromatin prevents mitotic chromosome segregation to cause hybrid lethality in Drosophila. *PLoS Biol.* **7**, e1000234 (2009).

[112] Maheshwari, S., Wang, J. & Barbash, D. A. Recurrent positive selection of the Drosophila hybrid incompatibility gene Hmr. *Mol. Biol. Evol.* **25**, 2421–2430 (2008).

[113] Tracy, C., Rio, J., Motiwale, M., Christensen, S. M. & Betran, E. Convergently recruited nuclear transport retrogenes are male biased in expression and evolving under positive selection in Drosophila. *Genetics* **184**, 1067–1076 (2010).

[114] Greil, F., de Wit, E., Bussemaker, H. J. & van Steensel, B. Hp1 controls genomic targeting of four novel heterochromatin proteins in Drosophila. *EMBO J.* **26**, 741–751 (2007).

[115] Sweigart, A. L. Simple Y-autosomal incompatibilities cause hybrid male sterility in reciprocal crosses between Drosophila virilis and D. americana. *Genetics* **184**, 779–787 (2010).

[116] Woodruff, G. C., Eke, O., Baird, S. E., Felix, M. A. & Haag, E. S. Insights into species divergence and the evolution of hermaphroditism from fertile interspecies hybrids of Caenorhabditis nematodes. *Genetics* **186**, 997–1012 (2010).

[117] Good, J. M., Vanderpool, D., Smith, K. L. & Nachman, M. W. Extraordinary sequence divergence at Tsga8, an X-linked gene involved in mouse spermiogenesis. *Mol. Biol. Evol.* **28**, 1675–1686 (2011).

[118] Bomblies, K. *et al.* Autoimmune response as a mechanism for a Dobzhansky-Muller-type incompatibility syndrome in plants. *PLoS Biol.* **5**, e236 (2007).

[119] Ellison, C. K. & Burton, R. S. Interpopulation hybrid breakdown maps to the mitochondrial genome. *Evolution* **62**, 631–638 (2008).

[120] Heck, J. A. *et al.* Negative epistasis between natural variants of the saccharomyces cerevisiae mlh1 and pms1 genes results in a defect in mismatch repair. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 3256–3261 (2006).

[121] Seidel, H. S., Rockman, M. V. & Kruglyak, L. Widespread genetic incompatibility in C. elegans maintained by balancing selection. *Science* **319**, 589–594 (2008).

[122] Lachance, J. & True, J. R. X-autosome incompatibilities in Drosophila melanogaster: tests of Haldane's rule and geographic patterns within species. *Evolution* **64**, 3035–3046 (2010).

[123] Chang, A. S. & Noor, M. A. Epistasis modifies the dominance of loci caus-

ing hybrid male sterility in the Drosophila pseudoobscura species group. *Evolution* **64**, 253–260 (2010).

[124] Coyne, J. A. Genetics and speciation. *Nature* **355**, 511–515 (1992).

[125] Wittbrodt, J. *et al.* Novel putative receptor tyrosine kinase encoded by the melanoma-inducing Tu locus in Xiphophorus. *Nature* **341**, 415–421 (1989).

[126] Barbash, D. A. & Ashburner, M. A novel system of fertility rescue in Drosophila hybrids reveals a link between hybrid lethality and female sterility. *Genetics* **163**, 217–226 (2003).

[127] Barbash, D. A., Awadalla, P. & Tarone, A. M. Functional divergence caused by ancient positive selection of a Drosophila hybrid incompatibility locus. *PLoS Biol.* **2**, e142 (2004).

[128] Sturtevant, A. H. Genetic studies of Drosophila simulans. I. introduction. hybrids with Drosophila melanogaster. *Genetics* **5**, 488–500 (1920).

[129] Lachaise, D., David, J. R., Lemeunier, F., Tsacas, L. & Ashburner, M. The reproductive relationships of Drosophila sechellia with D. mauritiana, D. simulans, and D. melanogaster from the afrotropical region. *Evolution* **40**, pp. 262–271 (1986). URL `http://www.jstor.org/stable/2408806`.

[130] Hutter, P. & Ashburner, M. Genetic rescue of inviable hybrids between Drosophila melanogaster and its sibling species. *Nature* **327**, 331–333 (1987).

[131] Clark, A. G. *et al.* Evolution of genes and genomes on the Drosophila phylogeny. *Nature* **450**, 203–218 (2007).

[132] Birney, E., Clamp, M. & Durbin, R. GeneWise and genomewise. *Genome Res.* **14**, 988–995 (2004).

[133] Bachtrog, D., Thornton, K., Clark, A. & Andolfatto, P. Extensive introgression of mitochondrial DNA relative to nuclear genes in the Drosophila yakuba species group. *Evolution* **60**, 292–302 (2006).

[134] Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680 (1994).

[135] Smit, A. F. A., Hubley, R. & Green, P. RepeatMasker Open-3.0. `http://www.repeatmasker.org` (1996-2010).

[136] Rozas, J., Sanchez-DelBarrio, J. C., Messeguer, X. & Rozas, R. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics (Oxford, England)* **19**, 2496–2497 (2003).

[137] DuMont, V. B., Fay, J. C., Calabrese, P. P. & Aquadro, C. F. DNA variability and divergence at the notch locus in Drosophila melanogaster and D. simulans: a case of accelerated synonymous site divergence. *Genetics* **167**, 171–185 (2004).

[138] Kumar, S., Tamura, K. & Nei, M. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief. Bioinform* **5**, 150–163 (2004).

[139] Yang, Z. PAML: a program package for phylogenetic analysis by maximum likelihood. *CABIOS* **13**, 555–556 (1997).

[140] Yang, Z. & Nielsen, R. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol. Biol. Evol.* **19**, 908–917 (2002).

[141] Cuff, J. A. & Barton, G. J. Application of multiple sequence alignment profiles to improve protein secondary structure prediction. *Proteins* **40**, 502–511 (2000).

[142] England, B. P., Admon, A. & Tjian, R. Cloning of Drosophila transcription factor Adf-1 reveals homology to Myb oncoproteins. *Proc. Natl. Acad. Sci. U. S. A.* **89**, 683–687 (1992).

[143] Cutler, G., Perry, K. M. & Tjian, R. Adf-1 is a nonmodular transcription factor that contains a TAF-binding Myb-like motif. *Mol. Cell. Biol.* **18**, 2252–2261 (1998).

[144] Aasland, R., Stewart, A. F. & Gibson, T. The SANT domain: a putative DNA-binding domain in the SWI-SNF and ADA complexes, the transcriptional co-repressor N-CoR and TFIIIB. *Trends Biochem. Sci.* **21**, 87–88 (1996).

[145] Bhaskar, V. & Courey, A. J. The MADF-BESS domain factor Dip3 potentiates synergistic activation by Dorsal and Twist. *Gene* **299**, 173–184 (2002).

[146] Boyer, L. A., Latek, R. R. & Peterson, C. L. The SANT domain: a unique histone-tail-binding module? *Nat. Rev. Mol. Cell Biol.* **5**, 158–163 (2004).

[147] Alba, M. M. & Guigo, R. Comparative analysis of amino acid repeats in rodents and humans. *Genome Res.* **14**, 549–554 (2004).

[148] Hancock, J. M. & Simon, M. Simple sequence repeats in proteins and their significance for network evolution. *Gene* **345**, 113–118 (2005).

[149] Ting, C. T., Tsaur, S. C. & Wu, C. I. The phylogeny of closely related species as revealed by the genealogy of a speciation gene, Odysseus. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 5313–5316 (2000).

[150] Kliman, R. M. *et al.* The population genetics of the origin and divergence of the Drosophila simulans complex species. *Genetics* **156**, 1913–1931 (2000).

[151] McDonald, J. H. & Kreitman, M. Adaptive protein evolution at the Adh locus in Drosophila. *Nature* **351**, 652–654 (1991).

[152] Wen, S. Y., Shimada, K., Kawai, K. & Toda, M. J. Strong purifying selection on the Odysseus gene in two clades of sibling species of the Drosophila montium species subgroup. *J. Mol. Evol.* **62**, 659–662 (2006).

[153] Begun, D. J. & Whitley, P. Adaptive evolution of relish, a Drosophila NF-kappaB/IkappaB protein. *Genetics* **154**, 1231–1238 (2000).

[154] Levine, M. T. & Begun, D. J. Comparative population genetics of the immunity gene, Relish: is adaptive evolution idiosyncratic? *PloS One* **2**, e442 (2007).

[155] Llopart, A. & Comeron, J. M. Recurrent events of positive selection in independent Drosophila lineages at the spermatogenesis gene roughex. *Genetics* **179**, 1009–1020 (2008).

[156] Malik, H. S. & Henikoff, S. Positive selection of Iris, a retroviral envelope-derived host gene in Drosophila melanogaster. *PLoS Genet.* **1**, e44 (2005).

[157] Wu, C. I. The genic view of the process of speciation. *J. Evol. Biol.* **14**, 851–865 (2001).

[158] Lipsick, J. S. One billion years of Myb. *Oncogene* **13**, 223–235 (1996).

[159] Oh, I. H. & Reddy, E. P. The myb gene family in cell growth, differentiation and apoptosis. *Oncogene* **18**, 3017–3033 (1999).

[160] Grune, T. *et al.* Crystal structure and functional analysis of a nucleosome recognition module of the remodeling factor ISWI. *Mol. Cell* **12**, 449–460 (2003).

[161] Fondon, J. W. r. & Garner, H. R. Molecular origins of rapid and continuous morphological evolution. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 18058–18063 (2004).

[162] Costa, R., Peixoto, A. A., Barbujani, G. & Kyriacou, C. P. A latitudinal cline in a Drosophila clock gene. *Proc. Biol. Sci.* **250**, 43–49 (1992).

[163] Rosato, E., Peixoto, A. A., Costa, R. & Kyriacou, C. P. Linkage disequilibrium, mutational analysis and natural selection in the repetitive region of the clock gene, period, in Drosophila melanogaster. *Genet. Res.* **69**, 89–99 (1997).

[164] Orr, H. A. The genetic basis of reproductive isolation: insights from Drosophila. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 6522–6526 (2005).

[165] Presgraves, D. C. The molecular evolutionary basis of species formation. *Nat Rev Genet* **11**, 175–180 (2010).

[166] Watanabe, T. K. A gene that rescues the lethal hybrids between Drosophila melanogaster and Drosophila simulans. *Jpn. J. Genet.* **54**, 325–331 (1979).

[167] Groth, A. C., Fish, M., Nusse, R. & Calos, M. P. Construction of transgenic Drosophila by using the site-specific integrase from phage phiC31. *Genetics* **166**, 1775–1782 (2004).

[168] Bischof, J., Maeda, R. K., Hediger, M., Karch, F. & Basler, K. An optimized transgenesis system for Drosophila using germ-line-specific phiC31 integrases. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 3312–3317 (2007).

[169] Venken, K. J., He, Y., Hoskins, R. A. & Bellen, H. J. P[acman]: a BAC transgenic platform for targeted insertion of large DNA fragments in D. melanogaster. *Science* **314**, 1747–1751 (2006).

[170] Fiumera, A. C., Dumont, B. L. & Clark, A. G. Sperm competitive ability in

Drosophila melanogaster associated with variation in male reproductive proteins. *Genetics* **169**, 243–257 (2005).

[171] Abramoff, M. D., Magelhaes, P. J. & Ram, S. J. Image processing with imagej. *Biophotonics International* **11**, 36–42 (2004).

[172] Kellum, R., Raff, J. W. & Alberts, B. M. Heterochromatin protein 1 distribution during development and during the cell cycle in Drosophila embryos. *J. Cell Sci.* **108 (Pt 4)**, 1407–1418 (1995).

[173] Ebert, A., Lein, S., Schotta, G. & Reuter, G. Histone modification and the control of heterochromatic gene silencing in Drosophila. *Chromosome Res.* **14**, 377–392 (2006).

[174] Hsieh, T. & Brutlag, D. Sequence and sequence variation within the 1.688 g/cm3 satellite DNA of Drosophila melanogaster. *J. Mol. Biol.* **135**, 465–481 (1979).

[175] Kas, E. & Laemmli, U. K. In vivo topoisomerase ii cleavage of the Drosophila histone and satellite iii repeats: DNA sequence and structural characteristics. *EMBO J.* **11**, 705–716 (1992).

[176] Lohe, A. R., Hilliker, A. J. & Roberts, P. A. Mapping simple repeated DNA sequences in heterochromatin of Drosophila melanogaster. *Genetics* **134**, 1149–1174 (1993).

[177] Abad, J. P. *et al.* Dodeca satellite: a conserved G+C-rich satellite from the centromeric heterochromatin of Drosophila melanogaster. *Proc. Natl. Acad. Sci. U. S. A.* **89**, 4663–4667 (1992).

[178] Carmena, M., Abad, J. P., Villasante, A. & Gonzalez, C. The Drosophila melanogaster dodecasatellite sequence is closely linked to the centromere and can form connections between sister chromatids during mitosis. *J. Cell Sci.* **105 (Pt 1)**, 41–50 (1993).

[179] Bonaccorsi, S. & Lohe, A. Fine mapping of satellite DNA sequences along the Y chromosome of Drosophila melanogaster: Relationships between satellite sequences and fertility factors. *Genetics* **129**, 177–189 (1991).

[180] Peng, J. C. & Karpen, G. H. H3K9 methylation and RNA interference regulate nucleolar organization and repeated DNA stability. *Nat. Cell Biol.* **9**, 25–35 (2007).

[181] Landry, C. R. *et al.* Compensatory cis-trans evolution and the dysregulation of gene expression in interspecific hybrids of Drosophila. *Genetics* **171**, 1813–1822 (2005).

[182] Prigent, S. R., Matsubayashi, H. & Yamamoto, M. T. Transgenic Drosophila simulans strains prove the identity of the speciation gene Lethal hybrid rescue. *Genes Genet. Syst.* **84**, 353–360 (2009).

[183] Orr, H. A. & Presgraves, D. C. Speciation by postzygotic isolation: forces, genes and molecules. *BioEssays* **22**, 1085–1094 (2000).

[184] Wu, C. I. & Ting, C. T. Genes and speciation. *Nat. Rev. Genet.* **5**, 114–122 (2004).

[185] Nolte, V., Weigel, D. & Schlotterer, C. The impact of shared ancestral variation on hybrid male lethality: a 16 codon indel in the Drosophila simulans Lhr gene. *J. Evol. Biol.* **21**, 551–555 (2008).

[186] Zhang, Z., Huang, J., Wang, Z., Wang, L. & Gao, P. Impact of indels on the flanking regions in structural domains. *Mol. Biol. Evol.* **28**, 291–301 (2011).

[187] Hormozdiari, F. *et al.* The effect of insertions and deletions on wirings in protein-protein interaction networks: a large-scale study. *J. Comput. Biol.* **16**, 159–167 (2009).

[188] Podlaha, O. & Zhang, J. Positive selection on protein-length in the evolution of a primate sperm ion channel. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 12241–12246 (2003).

[189] Podlaha, O., Webb, D. M., Tucker, P. K. & Zhang, J. Positive selection for indel substitutions in the rodent sperm protein catsper1. *Mol. Biol. Evol.* **22**, 1845–1852 (2005).

[190] Carroll, S. B. Endless forms: the evolution of gene regulation and morphological diversity. *Cell* **101**, 577–580 (2000).

[191] Hoekstra, H. E. & Coyne, J. A. The locus of evolution: evo devo and the genetics of adaptation. *Evolution* **61**, 995–1016 (2007).

[192] Mullaney, J. M., Mills, R. E., Pittard, W. S. & Devine, S. E. Small insertions and deletions (INDELs) in human genomes. *Hum. Mol. Genet.* **19**, 131–6 (2010).

[193] Wang, E. T. *et al.* Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470–476 (2008).

[194] Mills, R. E. *et al.* Natural genetic variation caused by small insertions and deletions in the human genome. *Genome Res.* **21**, 830–839 (2011).

[195] Verschure, P. J. *et al.* In vivo HP1 targeting causes large-scale chromatin condensation and enhanced histone lysine methylation. *Mol. Cell. Biol.* **25**, 4552–4564 (2005).

[196] Janssen, S., Durussel, T. & Laemmli, U. K. Chromatin opening of DNA satellites by targeted sequence-specific drugs. *Mol. Cell* **6**, 999–1011 (2000).

[197] Bhadra, M. P., Bhadra, U. & Birchler, J. A. Misregulation of sex-lethal and disruption of male-specific lethal complex localization in Drosophila species hybrids. *Genetics* **174**, 1151–1159 (2006).

[198] Chatterjee, R. N., Chatterjee, P., Pal, A. & Pal-Bhadra, M. Drosophila simulans Lethal hybrid rescue mutation (Lhr) rescues inviable hybrids by restoring X chromosomal dosage compensation and causes fluctuating asymmetry of development. *J. Genet.* **86**, 203–215 (2007).

[199] Li, H. *et al.* Cooperative and antagonistic contributions of two heterochromatin proteins to transcriptional regulation of the Drosophila sex determination decision. *PLoS Genet.* **7**, e1002122 (2011).